THE
PHYSICAL REVIEW
A journal of experimental and theoretical physics established by E. L. Nichols in 1893

SECOND SERIES, VOL. 67, NOS. 3 AND 4 FEBRUARY 1 AND 15, 1945

Measurement of the Specific Ionization of Fast Mesotrons with an Ionization Chamber and a Linear Amplifier*

W. CRAWFORD DUNLAP, JR.**
Department of Physics, University of California, Berkeley, California
(Received October 30, 1944)

The ionization of single cosmic-ray particles capable of penetrating 12 cm of lead has been measured with an ionization chamber and a linear amplifier. Measurements made at 14.7 atmospheres with argon indicate an average specific ionization of 71 ion pairs per cm at N.T.P. The results apply on the whole to mesotrons with an energy greater than 2×10^8 ev. The specific ionization of maximum frequency of occurrence is approximately 67 ion pairs per cm in argon. The average value for the ionization obtained with the ionization chamber at 14.7 atmospheres appears to be about 10 percent less than the total ionization. The average value obtained corresponds to the value 48 ion pairs per cm in air at N.T.P. as measured by a method that includes only collisions involving 10^4 ev or less. This result is in good agreement with the cloud-chamber results of Corson and Brode and of Hazen.

I. INTRODUCTION

ALTHOUGH the ionization chamber has been used very extensively for the measurement of ionization *currents* and for the measurement of ionization pulses of large numbers of cosmic-ray particles such as form a shower or burst, the only investigator previous to the present work who has used the instrument for the measurement of the ionization of single cosmic-ray particles is W. F. G. Swann.[1] Swann used a vertical cylinder filled with argon to a few atmospheres pressure. No counter control was used, and all types of pulse were recorded. A rather elaborate analysis had to be made to obtain the probable distribution of path lengths inside the cylinder. In abstract form Swann[2] has also mentioned use of counter control with the ionization chamber.

The work reported in the present paper was begun with the purpose of obtaining a new and independent value for the specific ionization of fast mesotrons, and if possible, of developing a method for obtaining accurately the specific ionization of a single mesotron. The momentum could be measured at the same time by means of a cloud-chamber measurement of the curvature of the track in a magnetic field. Although both the ionization and momentum can be obtained from cloud-chamber photographs, the diffuse tracks needed for accurate droplet counts make determination of the momentum inaccurate if only a single cloud chamber is used.

* This paper is a condensation (together with added material) of a thesis presented in August, 1942 to the Graduate Division of the University of California in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

** Now at the Western Regional Research Laboratory, Albany, California.

[1] W. F. G. Swann, Phys. Rev. **43**, 961 (1933).

[2] W. F. G. Swann, Phys. Rev. **45**, 258 (1934).

II. APPARATUS

A. General Description

A schematic diagram of the apparatus is shown in Fig. 1. The pulse produced by the motion of

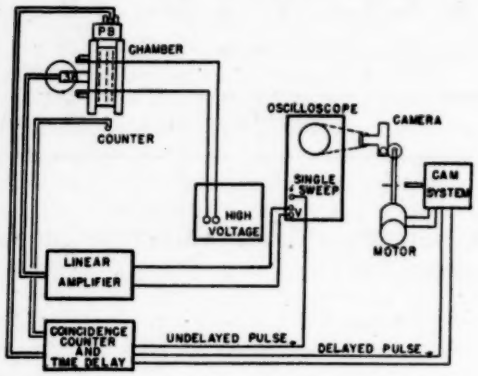

FIG. 1. Schematic diagram of apparatus. The chamber, counters, and 38 tube are mounted inside a shielded box, which is hung by a single piano wire from rubber cushioned supports to minimize microphonic effects.

the ions liberated by the primary mesotron or other particle was detected at the grid of the 38 tube. A high electrostatic field was used to collect the ions. The pulse was pre-amplified by the 38 tube and further amplified by the linear amplifier, which delivered the amplified pulse to the vertical deflection plates of the oscilloscope. The electron beam of the cathode-ray tube was deflected off the screen until a coincidence pulse caused the sweep-circuit thyratron to break down. The thyratron breakdown produced a single trace upon the screen. If the coincidence pulse was accompanied by an ionization pulse, as was ordinarily the case, the pulse appeared upon the screen at the start of the trace. Because of the difficulty of obtaining accurate visual observations of the length of the pulses, and because of the low counting rate—one in ten to twelve minutes—the apparatus was made self-recording and nearly automatic. In the arrangement finally adopted, the traces were photographed by means of a camera equipped with an automatic rewind, actuated by the pulse from the coincidence counter set. A cam system was used to cause the camera gear to be turned the proper fraction of a revolution after each pulse. The apparatus ran automatically for eight or nine hours on one loading of film.

B. The Ionization Chamber

A schematic diagram of the chamber in cross section is shown in Fig. 2. The chamber was cylindrical. The walls were designed to withstand a pressure of at least 70 atmospheres, although for the experiments reported here much lower pressures were used. Three plates were used in order to secure flexibility of operation. Thus, the two outer plates could be connected independently to high voltage sources of the same or opposite sign. In some cases it is desirable to use voltages of opposite sign so that for particles that produce equal ionizations in the two halves of the chamber no net effect is observed at the center electrode. The center collecting plate was insulated by use of amber that was coated with ceresin wax to reduce surface conductivity. The spark plugs used as leads for the high voltage plates were plugged with wax to make them gas-tight. Hard rubber insulators were found to be satisfactory supports for the high voltage plates at voltages up to 7500.

Microphonics arising especially from the vibrations of the plates were a serious problem throughout the course of the work. All three plates were made $\frac{3}{16}''$ thick to minimize these effects, and the whole apparatus was hung by a single piano wire from rubber-cushioned supports. These measures reduced the sensitivity of the system to vibration to the extent that only a very few pulse records had to be discarded because of the presence of microphonic disturbances.

C. The Electrical Circuits

The four-stage linear amplifier was of the negative-feedback type; the feedback was provided by use of un-bypassed cathode resistors and by two feed-back loops. The circuit was built up according to the design of Waddel[3] and differed from that of Waddel in only minor details. The amplification of the amplifier itself, including the 38 tube but not the oscilloscope amplifiers, was approximately 60,000. The calibration of the amplification system is discussed below.

The double-coincidence counter set was of standard design, and will not be described completely. The counter tubes were quenched by

[3] R. Waddel, Rev. Sci. Inst. **10**, 311 (1939).

means of the Neher-Harper[4] arrangement, and the coincidences were detected by use of the Rossi[5] coincidence circuit. The resolving time of the counter set was about 2×10^{-4} second. A time delay section employing an 885 thyratron was used to delay closing of the rewind-relay contacts until after the ionization pulse had been photographed. This measure was required because of the relative slowness of the ionization pulse compared to the coincidence pulse and because closing of the relay contacts during recording of the pulse would have caused sufficient sparking to obliterate the trace of the pulse. In order to operate the camera motor from the very short pulse from the coincidence counter, a standard self-holding telephone type relay with two energizing coils was used. The relay was closed by the coincidence pulse and remained closed until the holding current was stopped by the breaking of the relay contacts by the cam on the motor shaft.

The high electrostatic field required for the collection of the ions was obtained by use of a voltage doubler circuit that provided either plus or minus 2900 volts with respect to ground, or both. When it appeared desirable to use even higher voltages, a neon sign transformer was used in the rectifier set, and the circuit was changed so as to provide a single voltage variable from 0 to 7500 volts below ground. No results obtained by use of this voltage, however, are presented here. Since only small currents were drawn from the power supply of the high voltage unit, the little filtering required was provided by two sets of high resistances and one-microfarad condensers. There remained, however, some 60-cycle voltage in the background. This ripple apparently arose from pick-up of atmospheric fields, and was only partially removed by careful grounding of the shielding conductors.

III. CALIBRATION OF THE AMPLIFYING SYSTEM

The over-all amplitude and frequency characteristics of the amplifying system, including the oscilloscope amplifiers, were measured by use of alternating voltages obtained from a signal generator. From the frequency characteristic it was established that the amplification of the system was constant for frequencies corresponding to all collecting times used in the experiments.[6] From the amplitude characteristic it was verified that the length of the pulses was proportional to the voltage produced upon the grid of the 38 tube.

The amplitude and frequency characteristics mentioned above were only relative. The system was calibrated for what might be called the "absolute sensitivity" by use of a polonium alpha-particle source within the ionization chamber. The chief requirements for a good calibration are: (1) The pulses obtained must be of the same order of magnitude as the cosmic-ray pulses studied. (2) If possible, the particles should traverse the chamber in the same way as the cosmic-ray particles are expected to do. (3) The probable error of the result must be made small by observation of a large number of alpha-particle pulses. The method of calibration finally adopted involved use of an alpha-particle gun that confined the beam within a narrow solid angle in a direction parallel to the plates, and close to the voltage plate. The chamber was at a reduced pressure ($\sim$10 mm Hg) and the range of the particles was great enough to enable the alpha-particles to traverse the length of the plates much as a

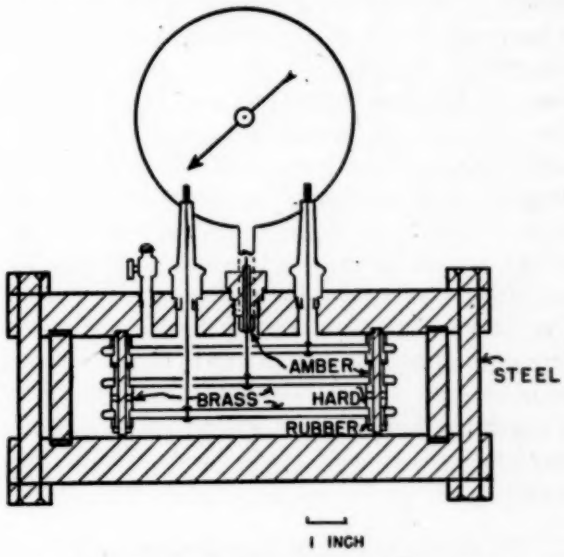

FIG. 2. Schematic cross-section diagram of the ionization chamber.

[4] V. H. Neher and W. W. Harper, Phys. Rev. **49**, 940 (1936).
[5] B. Rossi, Nature **125**, 636 (1930).
[6] Since these results were obtained with low impedance sources, they do not include an effect to be discussed below, namely, a variation of sensitivity with collecting time caused by the leakage of charge from the grid circuit of the 38 tube.

cosmic-ray particle would have done. By use of the ionization *versus* range data given by Rutherford, Chadwick, and Ellis,[7] the total number of ions expected in the sensitive region was calculated by graphical integration of the ionization. The resulting value of the sensitivity of the amplifying system was 22,300 ions per inch on the screen at the oscillograph gain setting used for the majority of the measurements. The standard deviation for this result is 2.0 percent.

One of the chief difficulties of the experiments was the fact that the time of collection of the ions, for high pressure, was not small compared to the time constant of the grid circuit of the 38 tube. This time constant is effectively the product of the leakage resistance from grid to ground and the capacity of the grid-collector system with respect to ground. The time constant for the apparatus used here was about 0.01 sec. The variation of pulse size with collecting time (or collecting field) was determined by use of the alpha-particle method mentioned above. By comparison of the resulting curve with the saturation curve for alpha-particles given in Jaffe's paper on columnar ionization,[8] it was found that in order to have the uncertainty caused by the collecting time effect less than 10 percent, the collecting time must be less than 0.005 sec. The collecting time must be less than 0.003 sec. for an uncertainty less than 5 percent. In addition, each value obtained at 27 atmospheres was increased by 25 percent, at 21.5 atmospheres by 12 percent, and at 14.7 atmospheres by 5 percent, to correct for loss of pulse size from the collecting time effect. The values for the collecting time used above were determined for the present calculations by use of the value 1.7 cm^2/volt sec. for the mobility of argon containing slight impurities of oxygen.[9] (Argon of 96 percent purity was used in this work.)

IV. SOME FURTHER CONSIDERATIONS AND ASSUMPTIONS

In the calculation of the specific ionization of the penetrating particles from the observed pulse sizes, the following assumptions were made: (1) Recombination can be neglected. Clay[10] has shown that for air in a shielded ionization chamber, with fields of the order of 2000 volts/cm and pressures of 15 to 25 atmospheres, the lack of saturation is only one or two percent. Since the pressures used in this work varied from 14.7 to 27 atmospheres, and the collecting field for most of the measurements was 1900 volts/cm, recombination can reasonably be neglected as a source of error. (2) The measured ionization is proportional to the pressure. Clay[11] has shown that the measured ionization in argon is proportional to the pressure, to within a few percent up to 60 atmospheres.

The path length used in the calculation of the probable values of the specific ionization is 17.0 cm. This length is not exactly that of the plates, but is a corrected value slightly greater than the length, 15.9 cm. The corrections take into account the effect of the fringing fields at the ends of the plates and also the fact that the Geiger counters subtended an appreciable solid angle. The counters were disposed so as to cover the entire region between two of the plates.

By use of the alpha-particle calibration and the above value of the probable path length, the "probable value of the specific ionization" was calculated, and it is equal to the observed ionization (corrected for collecting time effect) divided by the path length and by the pressure in atmospheres, reduced to standard conditions.

The expression "probable value of the specific ionization" should be used to denote the results, since it is not certain that a given pulse was produced by a single particle nor that the particle, assuming it to be single, traversed the chamber in the expected way. On the whole, however, the results can be ascribed to the action of single penetrating particles, presumably mesotrons. Starr[12] and others[13] have shown that the great majority, of the order of 90 percent, of double coincidences obtained with a few centimeters of lead between the counters is caused by single penetrating particles that do not produce showers. In the present experiments, the walls of the

[7] Rutherford, Chadwick, and Ellis, *Radiations from Radioactive Substances* (Cambridge University Press, 1930), p. 80.
[8] G. Jaffe, Ann. d. Physik **47**, 303 (1913).
[9] L. B. Loeb, *Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1934), p. 619.
[10] J. Clay and K. Oosthuizen, Physica **4**, 527 (1937).
[11] J. Clay and M. Kweiser, Physica **5**, 725 (1938).
[12] M. Starr, Phys. Rev. **53**, 6 (1938).
[13] S. H. Neddermeyer and C. D. Anderson, Phys. Rev. **51**, 884 (1937).

chamber, the block of lead, and the concrete roof above the apparatus are equivalent as far as ionization loss is concerned, to about 12 cm of lead. Using the Bethe-Bloch value of 1.5×10^6

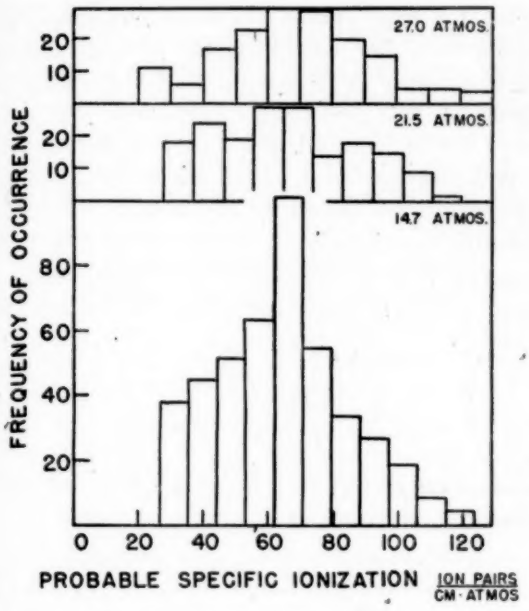

FIG. 3. Pulse-size-frequency distributions, plotted as block diagrams. The abscissa is the "probable specific ionization," that is, the specific ionization to be ascribed to the pulse assuming it to be produced by a single particle that traverses the chamber with path length 17.0 cm.

ev cm²/g, for the ionization loss of a fast mesotron at the minimum of the ionization *versus* energy curve, we see that the results apply on the whole to mesotrons with energies above 2×10^8 ev.

V. RESULTS AND DISCUSSION

Measurements were made at pressures of 27.0, 21.5, and 14.7 atmospheres of argon. Higher pressures were not used because of the limitations on the collecting time mentioned in Section III. About 1000 pulses were obtained in the form of photographic records, of which 700 were suitable for measurement. The block diagrams of Fig. 3 show the frequency of occurrence of certain specific ionization intervals of widths indicated. From the study previously mentioned of the variation of sensitivity of the amplifying system with time of collection, it was found that at 22 atmospheres there was an uncertainty of 20 percent in each individual result, 10 percent at 21.5 atmospheres, and about 5 percent at 14.7 atmospheres. In spite of this uncertainty, and of the previously mentioned corrections to the ionization values at the three pressures used, the maxima of the three distributions all appear to be in the same region. There is of course a pronounced broadening of the distributions at the higher pressures, which can be attributed to the collecting time effect.

The frequency distribution curve is considerably broader than would be expected if single particles having an energy spectrum such as that found by Blackett and obeying the Bethe-Bloch ionization law were passing through the chamber in the directions defined by the two Geiger-Müller counters.[14] In particular, it is to be noticed that there is a considerable number of pulses smaller than that most frequently occurring, as well as a large group of pulses longer than that most frequently occurring.

This broadening cannot be attributed to the "collecting-time effect," since the error in each result introduced by this effect at 14.7 atmospheres is only about 5 percent. There are, however, additional factors peculiar to the experimental method that may be responsible for the observed breadth of the curve. Some of these are: (1) Single penetrating particles may pass through the counters without traversing the full length of the sensitive region of the chamber to produce a small pulse. The coincidence pulse would most likely be produced in this case by scattering of the particle in the walls of the chamber or by production of a knock-on electron at the outer edge of the chamber. (2) Showers may trip the counters and one or more of the particles of the shower may pass obliquely through the sensitive region of the chamber to produce a small pulse. (3) Accidental coincidences may be caused by two independent cosmic-ray particles, one of which by chance may also pass obliquely through the sensitive region of the chamber to produce a small pulse. It is expected that showers are mainly responsible for the existence of both the very small pulses and the very large pulses observed.

The value taken for the specific ionization of maximum occurrence is 67 ion pairs per cm in argon at N.T.P. The *average* specific ionization is

[14] For information on the question of the constancy of specific ionization of cosmic ray particles with energy, see W. E. Hazen, Phys. Rev. **65**, 259 (1944).

also of interest. Using only the 14.7-atmosphere results we find by counting squares that the average specific ionization is 71 ion pairs per cm.

In order to obtain maximum accuracy in measurement of the most frequently occurring pulses, pulses with lengths more than about three times that most frequently occurring extended beyond the limits of the oscilloscope screen and were not measurable. By use of minor modifications of the photographic technique, the range of investigation can be modified so as to include any range of ionization, from that of single particles to that of importance in the study of showers and bursts.

If one uses Hopfield's assumption[15] that the ionization is proportional to the number of extranuclear electrons, then the specific ionization for air should be about 14.5/18 that for argon. For the average specific ionization in air from the results at 14.7 atmospheres, one obtains a value of approximately 60 ion pairs per cm.

Although a minimum pressure of 14.7 atmospheres was used in the chamber, it cannot be assumed that the average ionization obtained in this work represents the total ionization.[16] A secondary produced in the gas can impart only a limited energy to ionization and excitation within the sensitive region of the chamber. Assuming a constant ionization of 60 ion pairs per cm, and an average energy of 25 ev in argon for production of one ion pair,[17] we find the maximum energy of dissipation of a secondary to be 3.75×10^5 ev, if we assume that the secondary passes through the sensitive region in a straight line with path length 17.0 cm. This assumption is doubtful, however, since scattering of the particle would have the effect on the whole of shortening the path length by deflecting the particle into one of the plates. Secondaries with energies above 375 kev will be able to contribute only this same energy. Thus although it is rather difficult to calculate a precise upper limit to the secondary energies that are detected, for a rough calculation we take 375 kev as an effective upper limit and consider that secondaries with energies up to 375 kev are completely detected, and that secondaries with energies above 375 kev are completely undetected.

The total collision loss $\hat{\epsilon}_t$ in air is given by Rossi and Greisen[18] from the Bethe-Bloch theory to be 1.8×10^6 ev cm^2/g in air. At one atmosphere the total average energy loss is 2300 ev/cm. Rossi and Greisen also show that the average energy loss when only collisions up to 10^4 ev are considered is 28.5 percent less than the total, or $\hat{\epsilon}_{10^4}=1700$ ev/cm. We now wish to add to the latter figure the energy loss caused by collisions to secondaries between 10^4 and 3.75×10^5 ev. The probability of production of a secondary electron of energy E by a fast mesotron of high energy is given fairly accurately for air by the Rutherford formula[19]

$$\chi(E)dE=\frac{98dE}{E^2}\frac{\text{ev}}{\text{cm}}.$$

The average energy loss $\hat{\epsilon}_{E_1}^{E_2}$ caused by all such secondary electrons in the range E_1 to E_2 is

$$\hat{\epsilon}_{E_1}^{E_2}=\int_{E_1}^{E_2}E\chi(E)dE=98\int_{E_1}^{E_2}\frac{dE}{E}.$$

For $E_1=10^4$ ev, $E_2=3.75\times10^5$ ev,

$$\hat{\epsilon}_{10^4}^{3\times10^5}=98\ \ln\ 37.5\backsimeq360\ \text{ev/cm}.$$

Since the ionization chamber effectively measures $\hat{\epsilon}_{10^4}+\hat{\epsilon}_{10^4}^{3.75\times10^5}$, we find a fraction $2060/2300=90$ percent for the fraction of the total collision loss measured by the ionization chamber as used in the present work. Since the cloud-chamber result (for energies less than 10^4 ev) is 28 percent below the total ionization, the cloud-chamber result is about 20 percent less than the ionization chamber result. Thus for comparison of the present ionization chamber results with the cloud-chamber results, we should use the value $60-(0.2\times60)=48$ ions pairs per cm path in air at N.T.P.

Corson and Brode[20] found a minimum ionization of fast electrons in air of about 50 ion pairs per cm. Bagley[21] found an average ionization of

[15] J. J. Hopfield, Phys. Rev. **43**, 675 (1933).
[16] I am indebted to Dr. Wayne E. Hazen for pointing this out to me and for valuable assistance in connection with the point.
[17] See reference 7, p. 81.
[18] B. Rossi and K. Greisen, Rev. Mod. Phys. **13**, 240 (1941).
[19] B. Rossi and K. Greisen, Rev. Mod. Phys. **13**, 243 (1941). See also reference 14.
[20] D. Corson and R. B. Brode, Phys. Rev. **53**, 773 (1938).
[21] D. G. Bagley, Ph.D. Thesis, University of California (1941).

cosmic-ray particles in nitrogen of 43 ion pairs per cm. Hazen[22] has recently obtained a value of 50 ion pairs per cm for the average specific ionization in air of a large number of tracks of penetrating rays obtained at sea level and 100 feet underground.

It is of interest to compare the present results with those obtained by other workers using ionization chamber methods. Swann[23] obtained a value of 50 ion pairs per cm in argon by use of the uncontrolled ionization chamber having a thin wall and with little shielding around the chamber. Stuhlinger[24] in work with a proportional counter at one-half atmosphere, found a value for the specific ionization of the hard component in air of 35 ion pairs per cm.[25] Stuhlinger in this same work, and Swann[26] in work reported only in abstract form, have found a second maximum in the ionization curve. Indeed Stuhlinger gives a curve showing six or eight fairly well-defined maxima which he ascribes to various combinations of primary hard particles and softer secondary particles. The primaries are assumed from the position of the first maximum to have a specific ionization of about 35, and the secondaries then are found to have an ionization of about 50 ion pairs per cm, referred to air.

There is some evidence of such a second maximum in the 21.5-atmosphere results, but none in either the 27-atmosphere nor in the 14.7-atmosphere results. Since the low pressure results are the most reliable, it must be concluded that the present results indicate that there is only one maximum in the frequency distribution.

VI. CONCLUSION

The results obtained in this paper for the average value and for the most frequently occurring value of the specific ionization obtained under the described conditions are considered to be accurate to within eight or ten percent. The principal sources of error in these quantities are (1) the error of calibration of the sensitivity of the amplifying system, and (2) the fact that the pulses were classified into only a relatively few large groups. This classification was dependent upon the diffuseness of the trace, which in turn was determined by the background disturbances in the 38 tube.

In evaluating the question of the accuracy with which a *single* particle can be measured, we are probably upon safer ground if we examine the breadth of the experimental curve of Fig. 3 than if we rely upon estimates of the uncertainties that may be supposed to be introduced into each measurement. The curve has a half-breadth which appears to be some 25 or 30 percent of the most frequently occurring value. Although there is no reason to suppose that such a large error as 25 percent was inherent in the method of measurement of the ionization in the chamber, nevertheless, the presence of knock-on electrons from the walls or of pulses due to showers leads to essentially the same effect. Although use of the chamber in conjunction with a cloud chamber might remove the uncertainty as to whether a given pulse is associated with an air shower, there appears to be no way of being sure that a knock-on electron is not produced in the wall of the chamber to give a misleading impression of the specific ionization of the particle. Thus the usefulness of the method outlined in the present paper is greatest for work in which the most frequently occurring value, based upon as large a number of pulses as possible, can be used as the measure of the ionization.

VII. ACKNOWLEDGMENT

The present work was suggested by Professor R. B. Brode, to whom the author is also indebted for helpful suggestions and discussions, especially during the early phases of the investigation.

[22] W. E. Hazen, Phys. Rev. **65**, 259 (1944).
[23] W. F. G. Swann, Phys. Rev. **43**, 961 (1933). See also footnote, T. H. Johnson, Rev. Mod. Phys. **10**, 211 (1938).
[24] E. Stuhlinger, Zeits. f. Physik **108**, 444 (1942).
[25] Information has just been received of the publication of a paper "On the use of the linear amplifier for the measurement of the ionization of single particles" by S. A. Wytzes and G. J. van der Maas, Physica **10**, 419 (1943). No details, however, are available at the present writing.
[26] W. F. G. Swann, Phys. Rev. **61**, 393 (1942).

A Study of Time Variations in the Cosmic-Ray Directional Intensity Distribution*

MAX L. YEATER†
Department of Physics, Washington University, St. Louis, Missouri
(Received October 13, 1944)

Evidence of time variations in the cosmic-ray directional intensity distribution has been obtained, by a statistical analysis of directional intensity measurements carried out at St. Louis in 1943. The variations, occurring within periods of twelve to twenty-eight hours, seem most pronounced near zenith angles of $\pm 20°$ and $\pm 70°$ in the east–west plane, and are not found in the north–south plane. These observations would appear to be explained by the effect of terrestrial magnetic fluctuations in modifying the allowed cones of certain primary cosmic-ray particles, these particles having energies grouped near 0.2 Störmer and 0.6 Störmer. If this hypothesis is correct, the east–west positional symmetry of these variations constitutes new evidence, supporting previous evidence of Schremp, that positive and negative particles with the same e/m exist in the *primary* cosmic radiation. Earlier results of Ribner, Cooper, and Schremp, on the static fine structure pattern of cosmic-ray directional intensity, are also supported by the present data. The time variations in the intensity distribution reported here are variations in the fine structure pattern, but appear to be independent of the static prominences previously reported. The method for rapid exploration of the cosmic-ray directional intensity distribution is described. The multidirectional cosmic-ray telescope used is discussed in some detail.

I. INTRODUCTION

FROM theoretical considerations, based on the theory of the allowed cone[1] and the theory of atmospheric absorption of cosmic rays, Schremp predicted[2,3] that a fine structure pattern should exist in the cosmic-ray directional intensity distribution, and indicated[4] how a study of this pattern should provide information concerning the nature, energy spectra, and absorptive properties of the primary cosmic rays. The actual existence of the fine structure has been established by experiments of Ribner[5] and Cooper[6] in Missouri, and of Schremp and Baños[7] in Mexico City.

In Missouri the fine structure pattern was observed to have a circular symmetry about the zenith. This was interpreted as indicating the existence of lines or bands in the primary energy spectra, for particles too energetic to be observably influenced by the earth's magnetic field.[8] At Mexico City one circular ring of prominences was observed, corroborating the results obtained in Missouri. Other prominences in the Mexico City pattern failed to exhibit this circular symmetry, but were nearly mirror images of one another in the N–S plane. The latter provided evidence that the primary radiation contains positive and negative particles with the same e/m.[9]

Although these experiments dealt with the static characteristics of the fine structure pattern, the existence of time variations in the pattern was suspected.[10] In spite of the realization that new information might be revealed by the study of these time variations,[11] the matter could not be pursued at that time because of limitations in apparatus. The first attempt to explore such cosmic-ray fine structure time variations, with new apparatus, is discussed in this paper.

II. APPARATUS

A directional detecting system with a high counting rate combined with high angular resolution is needed. The cosmic-ray telescope used in this experiment was designed to measure radia-

* From a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at Washington University.

† Now at the Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts.

[1] M. S. Vallarta, J. Frank. Inst. **227**, 1 (1939) (summary); E. J. Schremp, Phys. Rev. **54**, 157 (1938); R. Albagli Hutner, Phys. Rev. **55**, 15 and 614 (1939).

[2] E. J. Schremp, Phys. Rev. **53**, 915A (1938).

[3] E. J. Schremp, Phys. Rev. **54**, 157 (1939).

[4] E. J. Schremp and H. S. Ribner, Rev. Mod. Phys. **11**, 149 (1939).

[5] H. S. Ribner, Phys. Rev. **56**, 1069 (1939).

[6] D. M. Cooper, Phys. Rev. **58**, 288 (1940).

[7] E. J. Schremp and Alfredo Baños, Jr., Phys. Rev. **59**, 614 (1941).

[8] See reference 4, p. 150.

[9] Cf. reference 7.

[10] Cf. references 4–6.

[11] Cf. references 4–6.

tion from eighteen directions simultaneously, using fairly large Geiger counters and large spacings between counters. The cross section for each of the directional systems is about 2.5 square feet. The dimensions are so adjusted that the geometrical angular aperture for any direction is 7° 40′ in the plane of the drum and 18° 18′ in the plane of the counter anode wires.

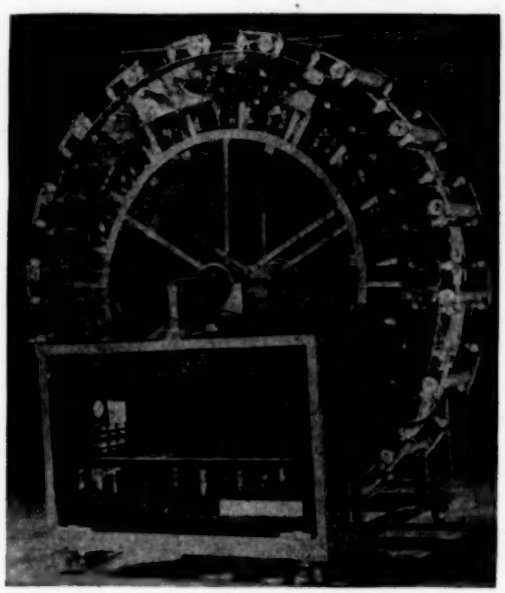

FIG. 1. Multidirectional cosmic-ray telescope. At the time this picture was taken only eighteen of the seventy-two Geiger counters were in position on the drum.

The use of this equipment, which is described in detail below, has indicated how certain improvements in apparatus would enable one to secure more complete information from this type of experiment. A summary of these suggested improvements in apparatus is included in Appendix III.

1. The Geiger-Müller Counters

The type of counter used is essentially that recommended by Trost,[12] consisting of a brass cathode, a tungsten anode, and a filling of argon plus organic vapor. A departure from conventional counter construction lies in the elimination of the usual glass envelope. The filling gas includes a mixture of alcohol and petroleum ether which minimizes temperature effects in the counting rate.

[12] Adolf Trost, Zeits. f. Physik **105**, 399 (1937).

Further details of the Geiger counters are given in Appendix I.

2. Mechanical Arrangement of Telescope

The multidirectional cosmic-ray telescope is shown in Fig. 1. The Geiger-Müller counters are mounted on the outer periphery of a drum six feet in diameter. One Neher-Pickering quenching circuit is provided for each set of counters; these circuits are mounted on a side panel which circles the drum on one side just below the counters. Shielded cables, encircling the drum in a layer beneath the counters, distribute pulses from the quenching circuits to the mixing circuits, which are below the cables. The high voltage power supplies are in a cabinet built into the base; the eighteen message registers which record the impulses for the eighteen directions are also in this cabinet. Slip rings and brushes link the circuits on the drum and in the cabinet. In the cabinet all the meters are on one panel; a movie camera is set up to photograph this panel automatically.

Changing the azimuth in which the plane of the drum lies can only be accomplished by moving the whole structure of the telescope. The zenith angle, however, can be changed automatically. The drive for this is provided by a $\frac{1}{8}$-hp, 1725-r.p.m. G.E. motor, with a 900:1 Boston "Reductor" and a 60:1 worm and gear. The automatic control circuit, for the drum and the camera, is in the cabinet with the power supplies.

3. Electric Circuits

a. Automatic Control Circuit

This circuit carries out these functions: It automatically starts the motor at periodic time intervals of twenty minutes, thirty minutes or one hour, and automatically stops the motor when the drum has rotated through a predetermined angular interval of one, three, or six degrees; it operates a message register which indicates the new position of the drum after the rotation, and automatically operates a camera to photograph the readings of the recording instruments at the end of the rotation of the drum. Provision is made for leaving the drum stationary and automatically photographing the instrument panel at intervals of twenty minutes, thirty minutes, or one hour.

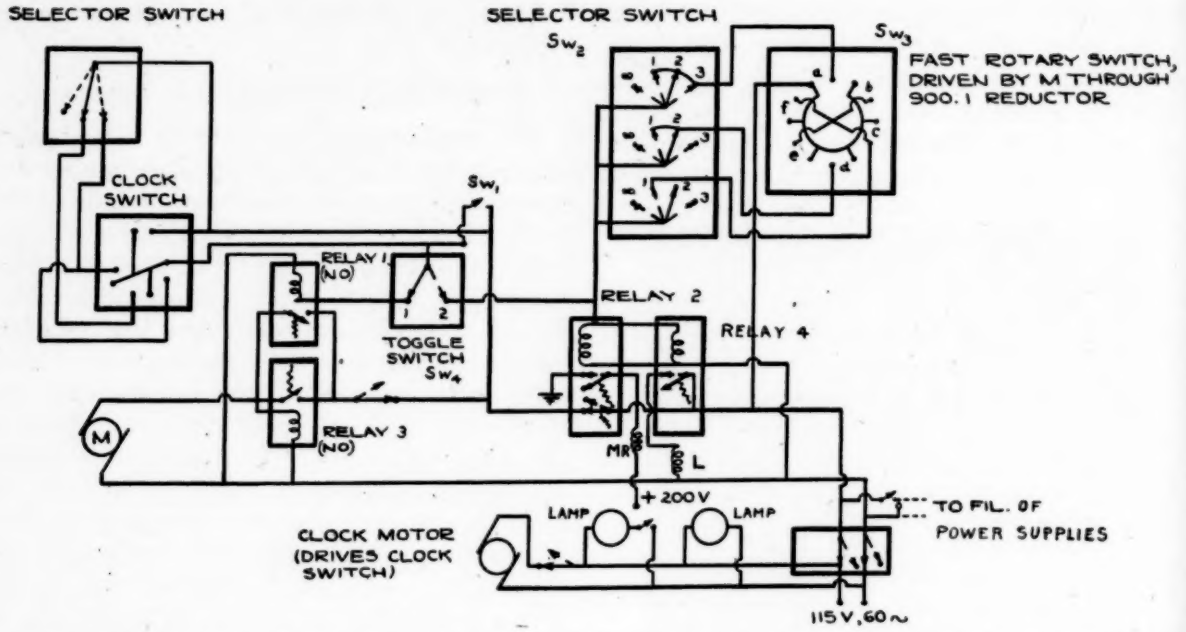

FIG. 2. Control circuit providing for automatic rotation of drum and photographing of meters in the multidirectional cosmic-ray telescope.

The essential features of this circuit, and the design of the special switch which makes this operation possible, have been described elsewhere;[13] a simplified circuit is given in this reference. Additional explanatory notes for the complete control circuit, which appears in Fig. 2, are given in Appendix IV.

b. Neher-Pickering Quenching Circuit

This is a conventional circuit; its operation is described elsewhere.[14] The constants are given in Fig. 3.

c. Mixing and Recording Circuit

This circuit, shown in Fig. 4, is a combination of the Rossi coincidence circuit[15] and the Johnson multivibrator circuit.[16] Its function is to accept pulses from five co-directional trains of counters (i.e., from the quenching circuits associated with these trains) and cause a message register to record a count whenever any one of these trains experiences a coincidence of cosmic-ray impulses in its members. The operation, briefly, is as follows: The 6C6 mixer tubes are normally conducting; whenever two negative pulses from counters on opposite sides of the drum come to the grids of any pair of the 6C6's, say A and B, the voltage across R_1' rises sharply and fires neon lamp Ne'. This causes the drop across R_2 to

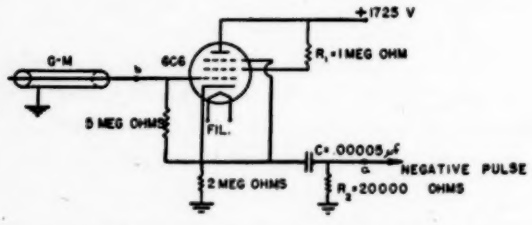

FIG. 3. Neher-Pickering quenching circuit.

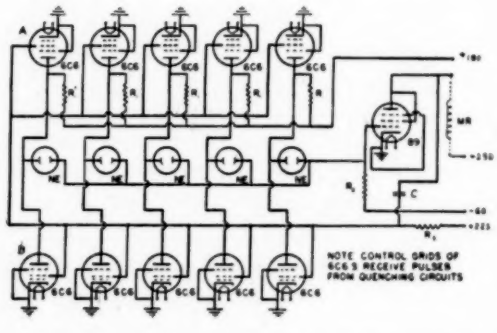

FIG. 4. Mixing and recording circuit for one direction; $R_1 = R_1' = 50{,}000$ ohms, $R_2 = 1$ megohm, $R_3 = 0.1$ megohm; Ne, ¼-watt neon lamps.

[13] M. L. Yeater, Rev. Sci. Inst. **14**, 146 (1943).
[14] A. L. Hughes, Am. Phys. T. **7**, 271 (1939).
[15] B. Rossi, Nature **125**, 636 (1930).
[16] T. H. Johnson, Rev. Sci. Inst. **9**, 221 (1938).

change so that the 89, a power tube, conducts. The duration of the conducting state in the 89 is increased by the feedback through C so that the message register coil MR in the plate circuit of the 89 can operate a mechanical numbering device.

Eighteen such circuits are used, one for each direction for which cosmic-ray intensities are measured. These are mounted on separate panels which encircle the drum; they are visible in Fig. 1, and have been identified in the description already given for the mechanical arrangement of the cosmic-ray telescope. The pulses received by these circuits from the quenching circuits must be sorted by a system of feeder cables, so that each mixer tube receives pulses only from the proper quenching circuit. This cable system is described in Appendix II.

d. Power Supplies

The regulated power supply for the plates of the 6C6 tubes in the coincidence circuits follows the circuit of Fig. 5. The screens of these tubes are provided with a similar supply which is also used for the negative voltage for the bias of the 89 power tube. The third supply, which operates the message registers through the power tube in the recording circuit, also uses the same circuit; two 2A3's are used in parallel, however, and heavier transformer and choke are used, because of the larger current.

The circuit shown in Fig. 6 provides the high voltage for the Geiger-Müller counters. In the type of experiment here involved, where the apparatus must run continuously for weeks or months, the first requirement for the high voltage supply is that it exhibit no drift in output voltage. It was found desirable to take these precautions:

(1) In order to avoid possible changes in the grid voltage for the 6J7, in the regulator circuit, the usual potentiometer was not used lest a defective contact arise at *a*, Fig. 6, where a changing potential is normally the means of controlling the output voltage.

(2) Batteries were used in the cathode circuit of the 6J7, in the regulator; neon lamps were first tried but were not reliable.

(3) The vacuum tubes are all mounted in the upright position. The voltage becomes noticeably erratic if this is not done.

When these precautions were taken the drift amounted to 1.5 percent at most; some of this may have been meter drift. No ripple greater than one volt could be detected in the output by the oscillograph.

As an additonal precaution, the input to all power supplies came from heavy duty Sola constant voltage transformers.

4. Light and Temperature Sensitivity

A serious difficulty arose from the effect of light on the sensitivity of the system. A periodic daily change in counting rate was observed. This effect was traced to the neon lamps in the multivibrator recording circuits. Coating them with Glyptal lowered the counting rate of the system by a factor of ten or more, but without changing the ratio of vertical to horizontal counting rate for one train (one directional pair of counting units). The recording circuits were very

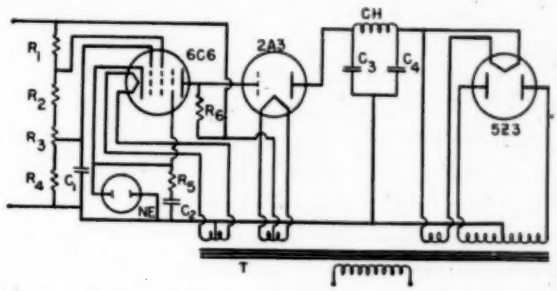

FIG. 5. Typical circuit for power supplies used for mixing and recording circuit.

$R_1 = 10^4$ ohms, 2 watts.
$R_2 = 25{,}000$ ohms, -2 watts.
$R_3 = 15{,}000$ ohms, wire wound potentiometer.
$R_4 = 5000$ ohms, 1 watt.
$R = 50{,}000$ ohms, 1 watt.
$R_6 = 250{,}000$ ohms, 1 watt.
$C_1 = 0.002$ f, tubular.
$C_2 = 0.01$ f, tubular.
$C_3 = C_4$ 16 f, electrolytic.
Ch = Thordarson choke, T-74C29.
T = Thordarson, T-13R08.
(For power tube plate power supply: Ch-T-74C29, T-13R09).

inefficient with neon lamps in darkness (normally they always received a little light at night from the light used in conjunction with the camera).

This difficulty was met by the direct method of putting extra lamps in the tent which housed the telescope, to provide continuous diffused illumination.

Ambient temperature changes for the cosmic-ray telescope were pronounced, as it was housed

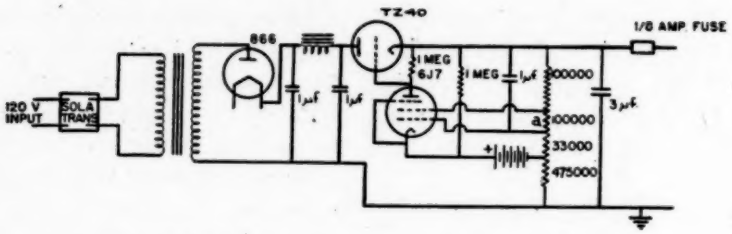

FIG. 6. High voltage power supply with regulator.

in a tent on the open roof of the Physics Building at Washington University. Temperature readings were recorded several times each day. No correlation between temperature and counting rate was discovered for temperatures above 20°F. Any temperature effects should affect all directions uniformly, and corrections were applied to the data to take account of possible aberrations of this type.

III. PROCEDURE

The system was operated continuously, except when the temperature was below 20°F. Every two to eight hours the drum was rotated by hand through two or two-and-one-half degrees; rotations were continuous in the same direction. The meter panel was photographed automatically every half-hour. Since the directional counting arrays are separated in angle by ten degrees, either five two-degree rotations or four two-and-one-half degree rotations provided data covering the complete 180-degree range in zenith angle. This complete coverage was accomplished in periods of from twelve to twenty-eight hours.

The advantages of this manual rotation of the telescope drum over automatic rotation are several, and a summary of these will serve to bring out more information concerning the experimental method used in this work: (1) The functioning of the telescope can be checked at the time of rotation of the drum, which is when unfortunate minor accidents are most likely to occur. (2) The routine calculations, exclusive of finding probable errors, are reduced by lengthening the time intervals during which the drum remains in one position; with automatic rotation this would have been one hour at most. (3) The total time interval for each zenith direction is broken up into subintervals by the automatic photographs, so that one can check the probable error of any point. (4) The number of counts per point on a complete intensity curve is statistically more significant than would be the case if the drum were at one position for only one hour or less. All of these advantages except (1) could be achieved with completely automatic operation of this apparatus if the control circuit were modified to permit the use of other angular intervals and other time intervals.

Whether operated manually or automatically, the present type of apparatus allows the angular interval between successive positions of the drum to be so chosen that each directional array is made to explore one "section"[17] of the zenith angle range. These sections can be joined to form one continuous curve, as described below.

Readings of temperature and atmospheric pressure were recorded several times daily while cosmic-ray observations were carried out.

The data were taken from the photographic record as soon as possible and later reduced as described below.

A few improvements in procedure which this work showed to be desirable are suggested in Appendix III.

IV. TREATMENT OF EXPERIMENTAL DATA

1. Corrections

It was considered advisable to make certain corrections. Since the western end of each section in the east–west curves, for example, begins at the same time, a changing counting rate caused by spurious effects common to the whole instrument (such as voltage changes or ambient temperature changes) could lead to spurious anomalies common to all sections. A good measure of the drift in counting rate is afforded by the total counting rate for all eighteen trains. In the

[17] The word "section" will henceforth, in this paper, be used in this sense, denoting a small portion of the zenith angle range.

absence of spurious effects this total counting rate should be nearly the same for every position traversed by the drum within a ten-degree section. The method of making the corrections, then, involves multiplying each recorded number of counts by such a factor that the total corrected number of counts for each position of the drum is the same and equal to the average total value calculated from the observed totals. The multiplying factors are just the ratios of the various observed total counting rates, one for each position of the drum, to their average value for the given ten degree section. The same factor applies to every one of the eighteen intensities recorded for each position of the drum.

These results of this method of applying corrections are to be noted: (1) Any common bending of the ten-degree sections, which might lead to spurious anomalies, will be straightened in this process. (2) Any directional effects of atmospheric or magnetic disturbances will show up only in the directions actually involved. For the anomalous change in one direction will be a small fraction of the total intensity in that direction, which in turn will certainly not be a large fraction of the total intensity for eighteen directions. (3) These corrections do not eliminate errors in relative intensity between any two sections, caused by different counting rates for different directional arrays. Although the sections can be adjusted to form one continuous curve, as described below, errors can accumulate in this curve in the process of bringing the separate ten-degree sections into coincidence at their terminal points. For a statistical error, say, in one terminal point may be introduced proportionally into every point of the adjacent section. This means that different sections, either before or after being merged into one continuous curve, can be compared with respect to shape but not with respect to intensity; and in particular, that east–west asymmetries cannot be determined from the present data.

2. Adjustment of Sections to Form Complete Curves

The system was rotated through four 2.5° intervals or five 2° intervals to secure data for one curve. Since the individual arrays are separated in angle by ten degrees, sections thus found have common end points. The adjustment of the sections to form a continuous curve involves multiplying each point of a section by that factor which makes one of its terminal points coincide with the terminal point of one adjacent section. The other adjacent section is then similarly matched to this section.

3. Probable Errors

Since automatic recording of meter readings occurred every half-hour, each time interval for one angle is effectively broken into subintervals. From these data one could compute the probable error of every point by usual methods.[18] Since an examination of the data indicated no excessive fluctuations, the labor of such calculations has been omitted. An estimate of the probable error of any point is given by $n^{\frac{1}{2}}$, where n is the number of counts. The statistical treatment described below, moreover, relieves the probable errors of final responsibility for the validity of the results.

4. Statistical Analysis

In order to find time variations in the directional intensity pattern, the daily curves are compared with the time-average of all such curves for the same azimuth. Pearson's criterion for goodness of fit[19] is used. Each ten-degree section of an individual curve is compared with the corresponding section of the time-average curve after the latter has been fitted as follows: Let n_θ be the number of counts per point, at the zenith angle θ, of the time-average curve. Let r_θ be the number of counts per point, at angle θ, of the individual curve. Then

$$\frac{n_\theta}{\sum_{\theta_i}^{\theta_i+10^\circ} n_\theta} \sum_{\theta_i}^{\theta_i+10^\circ} r_\theta = n_\theta'$$

$$(\theta_i = 0^\circ, 10^\circ, 20^\circ \cdots, \text{or } 2^\circ, 12^\circ, 22^\circ \cdots, \text{etc.})$$

gives the points for the time average section fitted to the individual section. The deviation δ_θ

[18] Cf. Gunnar Dahlberg, *Statistical Methods for Medical and Biological Students* (George Allen and Unwin, London, 1940), Chap. VIII; T. C. Fry, *Probability and Its Engineering Uses* (McGraw-Hill Book Company, Inc., New York, 1928).

[19] Discussed by Fry, reference 18, p. 285.

for any point of the individual curve is then

$$\delta_\theta = n_\theta' - r_\theta.$$

Note that

$$\sum_{\theta_i}^{\theta_i+10^\circ} n_\theta' = \sum_{\theta_i}^{\theta_i+10^\circ} r_\theta$$

and

$$\sum_{\theta_i}^{\theta_i+10^\circ} \delta_\theta = 0.$$

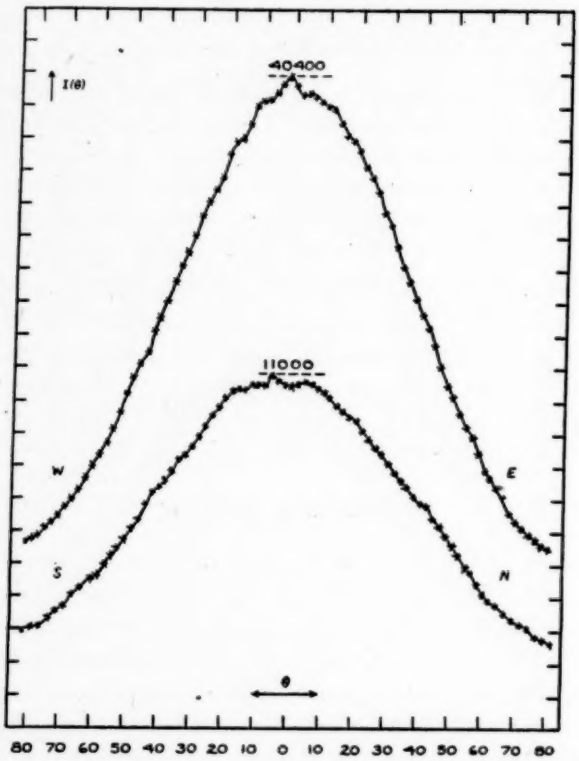

FIG. 7. Time average cosmic-ray directional intensity $I(\theta)$ as a function of zenith angle θ for the E–W azimuth (upper curve) and N–S azimuth (lower curve). For the E–W curve each vertical scale division represents 2000 counts; for the N–S curve each division represents 1000 counts. The maximum intensity for each curve is shown. The time interval covered by the E–W curve extends from 9:10 P.M., March 25, 1943 to 8:10 P.M., April 12, 1943; the time interval for the N–S curve is from 3:35 P.M., April 15, 1943 to 1:06 P.M., April 20, 1943. The average probable error is about 0.4 percent of the maximum intensity for the E–W curve and 0.7 percent of the maximum intensity for the N–S curve.

With the quantity

$$\chi^2 = \sum_{\theta_i}^{\theta_i+10^\circ} \frac{\delta_\theta^2}{n_\theta'},$$

one enters Pearson's table for goodness of fit criteria.[20] The probability number P given by the table is the probability that another run of the experiment would by chance yield a curve deviating from the standard of comparison as much or more than the curve at hand. In the present data each ten-degree section of a curve contains either five or six points; one therefore enters Pearson's table with n'^{20} equal to four or five, respectively, since a relationship between the points is implied by the method of fitting the time-average curve to the individual curve.

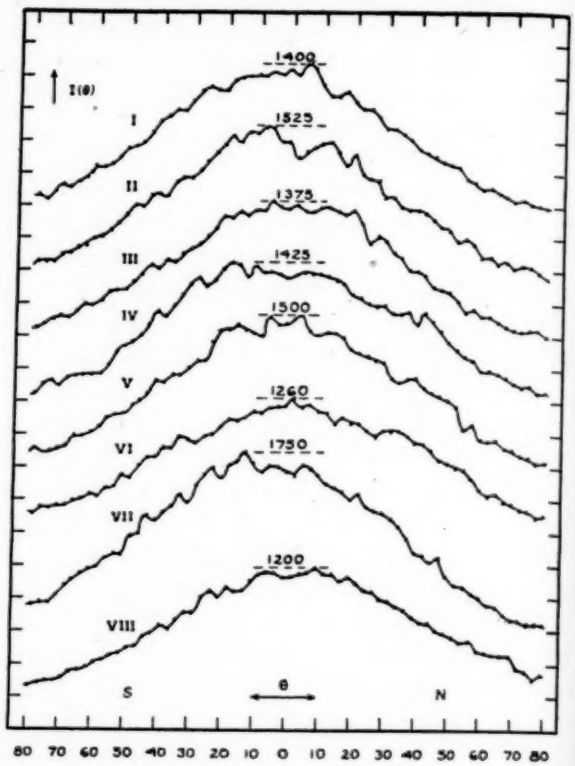

FIG. 8. Individual short-time directional intensity curves for the north–south azimuth. Each vertical scale unit represents 250 counts. The maximum intensity is shown for each curve. The average probable error is about 2 percent of the maximum intensity. The time intervals are as follows (all dates for 1943): I. 4–15, 3:35 P.M. to 4–16, 6:06 A.M. II. 4–16, 9:06 A.M. to 4–16, 7:06 P.M. III. 4–16, 11:36 P.M. to 4–17, 3:06 P.M. IV. 4–17, 3:36 P.M. to 4–18, 1:36 A.M. V. 4–18, 3:36 A.M. to 4–18, 7:36 P.M. VI. 4–18, 8:36 P.M. to 4–19, 8:36 A.M. VII. 4–19, 4:36 P.M. to 4:20, 2:36 A.M. VIII. 4–20, 3:06 A.M. to 4–20, 1:06 P.M.

V. RESULTS AND DISCUSSION

1. The Time-Average Directional Intensity Curves

For the east–west and north–south azimuths the cosmic-ray intensity $I(\theta)$ is plotted as a function of zenith angle θ in Figs. 7–9. The two

[20] Karl Pearson, *Tables for Statisticians and Biometricians* (Cambridge University Press, London, 1930), third edition, Part I, p. 26.

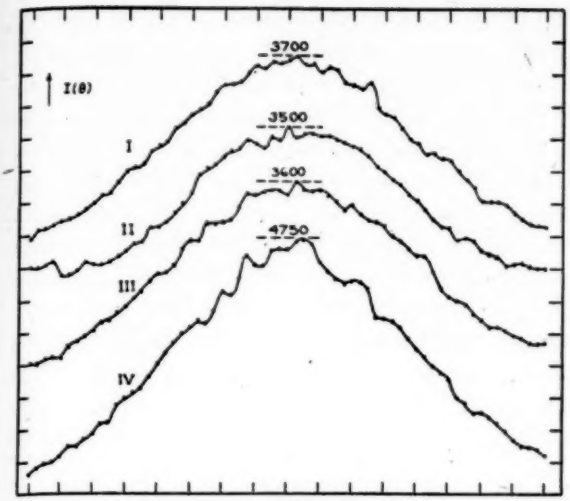

FIG. 9A.

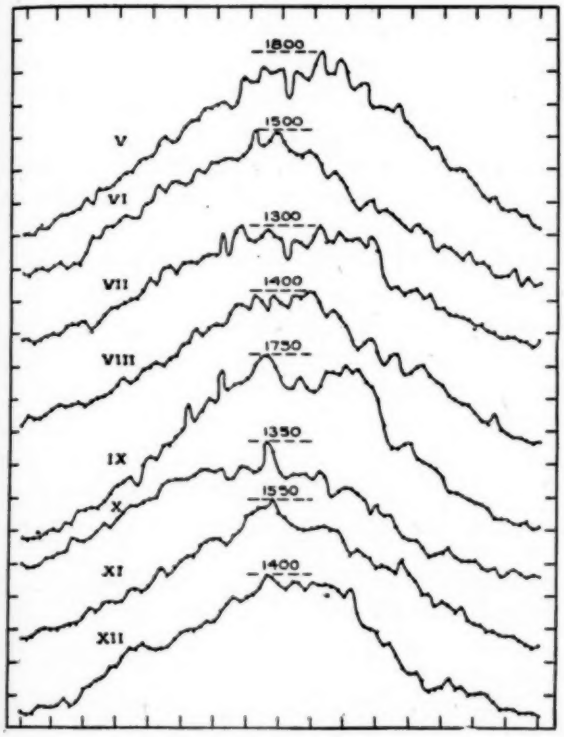

FIG. 9B.

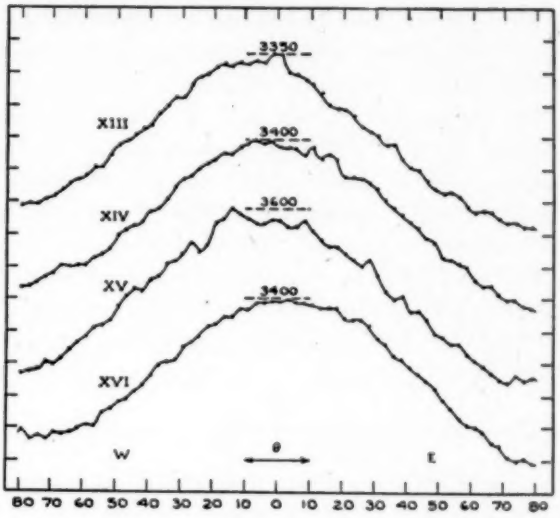

FIG. 9C.

FIGS. 9A–C. Individual directional intensity curves for the east–west azimuth. Each vertical scale unit represents 500 counts for curves I–IV and XIII–XVI, and 250 counts for curves V–XII. The maximum intensity for each curve is shown. The average probable error is about 1 percent of the maximum intensity for curves I–IV and XIII–XVI, and about 2 percent of the maximum intensity for curves V–XII. The time intervals are as follows (all dates for 1943): I. 3–25, 9:10 P.M. to 3–27, 1:10 A.M. II. 3–27, 2:40 A.M. to 3–28, 6:40 A.M. III. 3–29, 4:40 P.M. to 3–30, 7:10 P.M. IV. 3–30, 7:10 P.M. to 3–31, 9:10 P.M. Each of curves V, VII, IX, and XI covers 12 hours starting at 7 P.M. on 3–31, 4–1, 4–2, 4–3, respectively. Curves VI, VIII, X, and XII are for 12-hour periods starting at 7 A.M. on 4–1, 4–2, 4–3, 4–4, respectively. XIII. 4–7, 3:10 A.M. to 4–8, 8:10 A.M. XIV. 4–8, 11:40 A.M. to 4–10, 12:10 A.M. XV. 4–10, 9:40 A.M. to 4–11, 12:10 P.M. XVI. 4–11, 5:10 P.M. to 4–12, 8:10 P.M.

curves of Fig. 7 are summation, or time-average, curves representing all of the data recorded for the E–W and N–S azimuths, respectively; they are obtained by the superposition of the short-time curves of Figs. 8 and 9, which will be discussed in succeeding sections.

These time-average curves are similar in scope to the measurements made previously by Ribner[21] and Cooper[22] in Missouri, although more cosmic-ray counts are recorded here than in the earlier work. This permits a comparison of the present results with those previously reported.

In the present time-average curves prominences can be recognized near E 5°, E 10°, E 20°, W 5°, W 10°, W 20°, S 5°, S 10°, S 20°, and N 5°. At N 20°, where others have reported an anomaly, a prominence is just perceptible. Prominences at 40° in the north, south, east, and west are not strong but are sufficiently evident to bear out the earlier reports that anomalies do exist at these angles. In addition to the 40° anomaly, the data of Ribner and Cooper reveal prominences near 7° and 20° in the E and W azimuths (Ribner and Cooper) and in the N, S, NE, NW, SE, and SW azimuths (Cooper). The prominences in the

[21] See reference 5.
[22] See reference 6.

present curves near 5° and 10° are probably the same as appeared to be one near 7° from the more widely spaced points of the earlier work. The prominences at 20° in the N, S, E, and W appear to be the same for both the present and past data.

The present time-average curves, therefore, tend to confirm rather completely the existence and circular symmetry of the anomalies previously reported; they support the interpretation suggested at that time for this circular symmetry, namely, that lines or bands exist in the primary energy spectrum.[23]

2. The Daily Curves and Their Time Variations

Each of the twenty-four curves of Figs. 8 and 9 spans a period of from twelve to twenty-eight hours, and covers the full 180° of one azimuth at 2° or 2.5° intervals. An attempt was made to determine the existence and character of time variations in the directional cosmic-ray intensity merely from an inspection of these daily curves. This rather obvious step was found inadequate to provide any consistent interpretation of the data; selected combinations of these curves were similarly uninformative.

A better approach to the study of these curves is a section by section comparison of corresponding ten-degree angular intervals of the different curves;[24] this is best accomplished by a statistical analysis. The method adopted has already been described in Section IV, 4, and the results of this analysis will now be discussed.

For the east–west azimuth, Table I gives the probability values P describing the goodness of fit of each ten-degree section of the individual directional intensity curves with the time-average curve; Table II lists the corresponding values for the north–south azimuth. There are many entries for $P \lesssim 0.1$ and, in the E–W azimuth, even for $P \lesssim 0.01$. In determining whether these low values actually indicate significant time variations, three interrelated questions are to be dealt with: (1) A few low P values are to be expected just by chance; is the number of these values here so high as to imply the existence of real time variations? (2) What limiting value of P shall be considered useful for distinguishing abnormal sections of the individual curves, and hence for indicating the angular distribution of any moving prominences? (3) Does the angular distribution of the low P values indicate the absence of random instrumental errors?

The curves of Figs. 10 to 16 are helpful for a consideration of these questions, which will be taken up in order.

(1) In each case of Figs. 10 to 13, for the E–W azimuth, the horizontal "chance line" is appreciably below the actual distribution. This line becomes relatively lower, furthermore, as the limiting values of P are decreased from 0.2 to 0.01; the average ratios of chance expectation to actual number of sections are 0.6 for $P \lesssim 0.2$, 0.4 for $P \lesssim 0.1$, 0.3 for $P \lesssim 0.05$ and 0.17 for $P \lesssim 0.01$. The existence of significant time variations in the

[23] Cf. references 4, 6, and 7.

[24] For this purpose the original corrected data for the individual ten-degree sections are used, rather than the adjusted values which form the continuous curves of Figs. 8 and 9.

TABLE I. Probability values P describing the goodness of fit of the individual east–west curves with their time average.*

Curve / Angular section	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
W 90–80	0.11	0.78	0.50	0.70	0.44	0.87	0.50	0.83	0.22	0.90	0 70	0.53	0.88	0.30	0.92	0.86
80–70	.08	.56	.16	.13	.16	.48	.50	.28	.36	.57	.80	.35	.10	.90	.46	.02
70–60	.91	.10	.0011	.66	.49	.08	.62	.04	.018	.019	.70	.04	.68	.45	.05	.05
60–50	.54	.18	.22	.27	.35	.06	.0012	.97	.50	.70	.11	.46	.52	.02	.54	.08
50–40	.84	.38	.28	.80	.93	.65	.35	.40	.04	.13	.36	.85	.85	.70	.025	.33
40–30	.76	.42	.32	.32	.36	.13	.14	.54	.62	.46	.35	.74	.64	.45	.64	.52
30–20	.56	.011	.21	.30	.94	.65	.30		.25	.82	.88	.23	.86	.38	.0004	.50
20–10	.58	.56	.08	.018	.24	.75	.00008	.25	.03	.30	.40	.30	.10	.29	.001	.50
10–0	.58	.06	.74	.66	.00005	.30	.80	.58	.04	.10	.64		.48	.66	.30	.93
E 0–10	.10	.60	.06	.15		.57	.003	.10	.68	.86	.55	.96	.20	.98	.25	.52
10–20	.38	.45	.74	.56	.30			.08	.00004	.93	.83	.013	.08	.04	.82	.36
20–30	.00001	.84	.16	.04	.34	.10	.05			.035	.75	.12	.26	.21	.04	.33
30–40	.28	.92	.54	.20	.15	.40	.28	.047	.18		.12	.00004	.018	.80	.10	.56
40–50	.0014	.62	.035	.10	.72	.09	.40	.14	.49	.15	.36		.025	.97	.06	.88
50–60	.41	.12	.62	.05	.78	.17	.60	.42	.67	.047	.55	.07		.18	.76	.06
60–70	.0024	.19	.08	.88	.52	.12	.32	.0025	.53	.49	.85	.16	.74	.07	.10	.05
70–80	.21	.48	.84	.68	.044	.0012	.40	.05	.59	.96	.73	.64	.024	.30	.0003	.0002
80–90	.50	.72	.95	.49	.32	.75	.50	.20	.19	.90	.08	.37	.035	.33	.44	.08

* The missing entries for this table represent data thrown out because of instrumental errors.

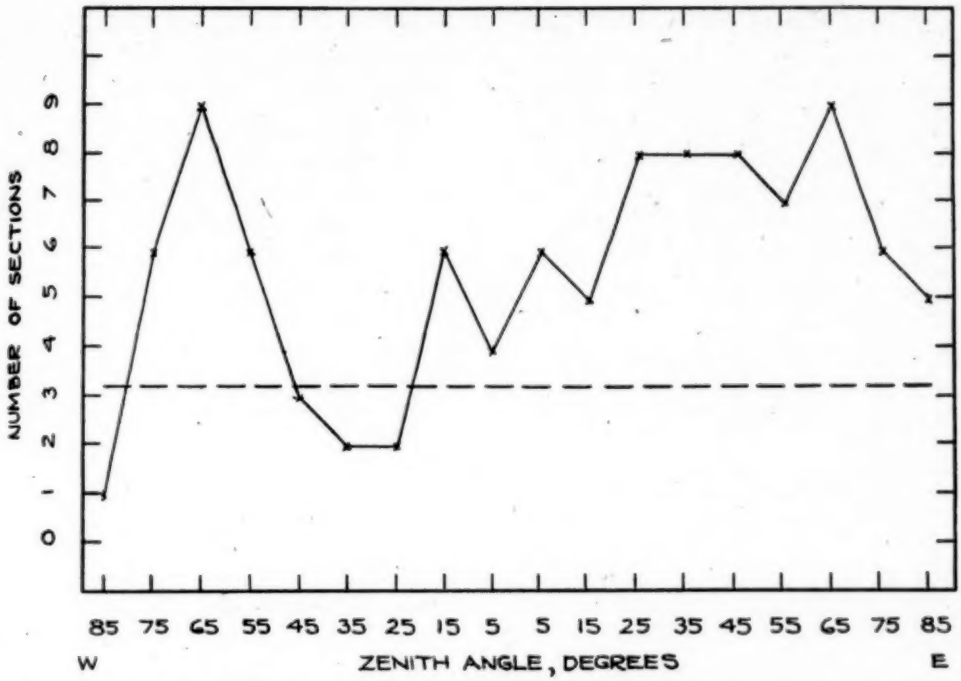

FIG. 10. Number of 10° sections of individual E–W directional intensity curves for which $P \leq 0.2$, as a function of zenith angle. The broken horizontal "chance line" shows the most probable number expected by chance for any 10° interval.

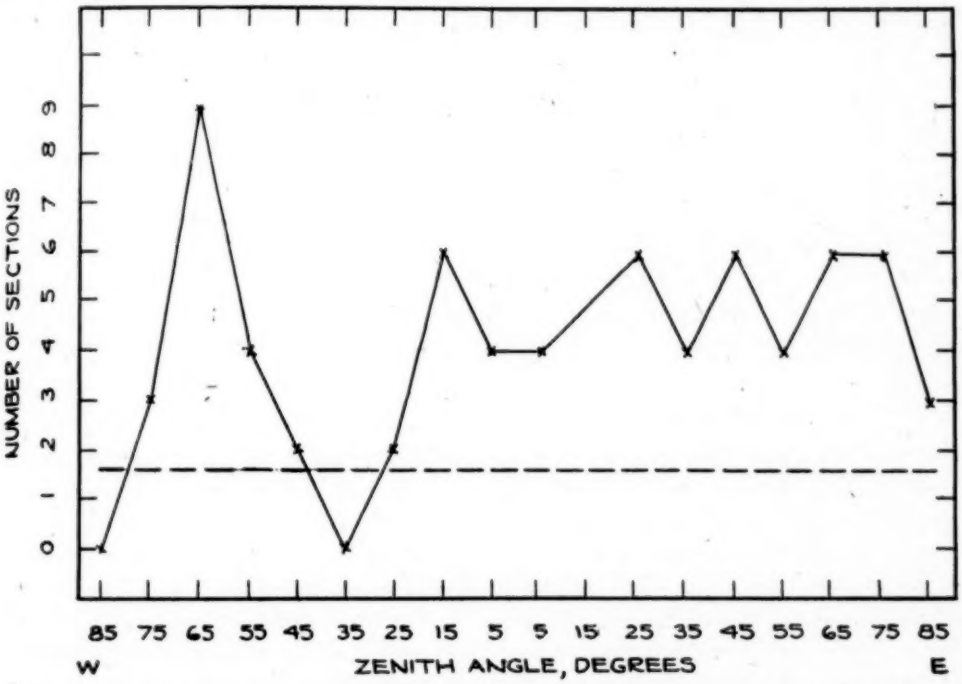

FIG. 11. Number of 10° sections of individual E–W directional intensity curves for which $P \leq 0.1$, as a function of zenith angle. The broken line shows the most probable number expected by chance for any 10° interval.

east–west directional intensity pattern therefore seems to be indicated, provided only that random instrumental errors can be shown to be absent.

The results for the north–south azimuth are

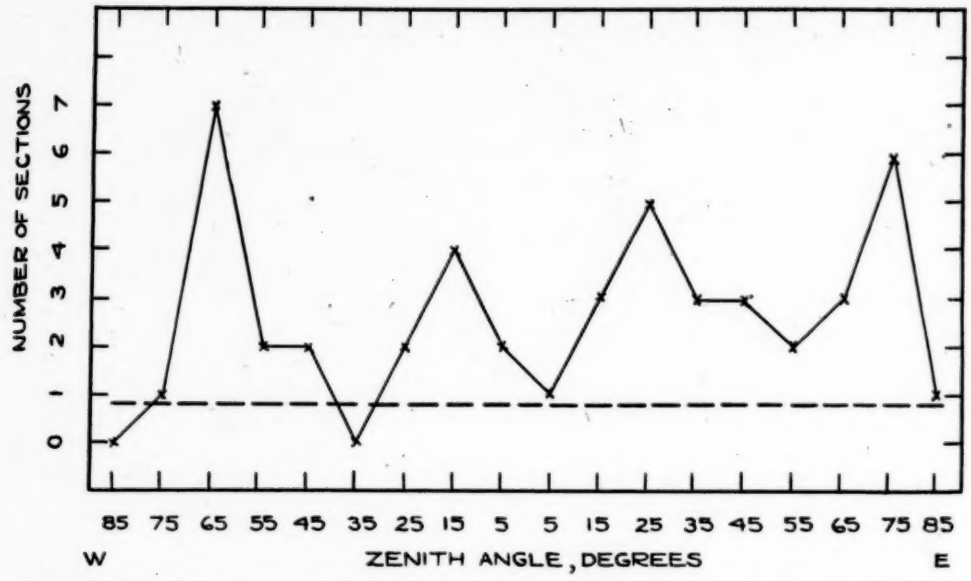

FIG. 12. Number of 10° sections of individual E–W directional intensity curves for which $P \lesssim 0.05$, as a function of zenith angle. The broken line shows the most probable number expected by chance for any 10° interval.

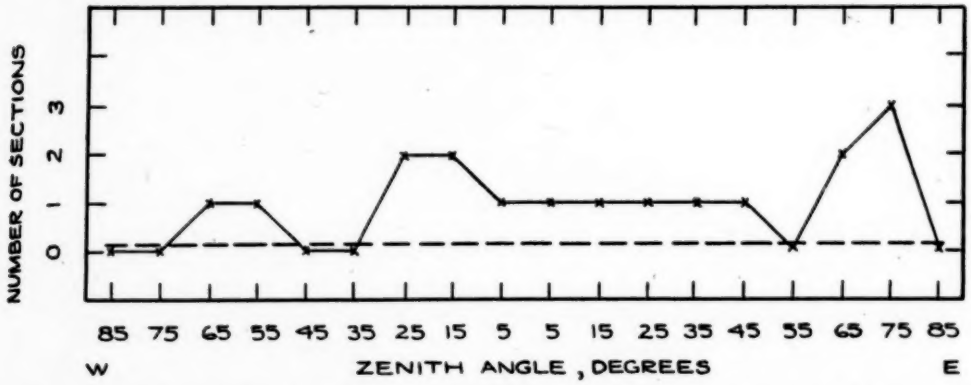

FIG. 13. Number of 10° sections of individual E–W directional intensity curves for which $P \lesssim 0.01$, as a function of zenith angle. The broken line shows the most probable number expected by chance for any 10° interval.

different. For $P \lesssim 0.2$ and $P \lesssim 0.1$ the number of sections is in no case appreciably greater than could be expected by chance (cf. Figs. 15 and 16). Only two sections exhibit a value less than 0.05, although seven would be the normal chance expectation; none has $P \lesssim 0.01$. This situation will be discussed further after questions (2) and (3) have been taken up. For the moment we merely point out this difference between the data for the N–S and E–W azimuths, a difference which implies that time variations of a type which can be revealed by Pearson's test exist in the E–W azimuth but not in the N–S azimuth.

(2) It is reasonable to use the plot of $P \lesssim 0.05$ as a guide in locating the angular distribution of time-varying sections of the pattern. This choice can be justified to some extent by the ratio of the number of sections expected by chance to have $P \lesssim 0.05$, to the number actually observed, for the E–W azimuth (cf. Fig. 12). The average value of this ratio is 0.3 and even lower values exist for several angles, the minimum being 0.11. The corresponding ratio (average value equals 0.17) for $P \lesssim 0.01$ of course does not preclude 0.01 as an upper limit of P, but the observed number of sections for this case is too small to be useful.

(3) The angular distribution for $P \lesssim 0.05$ in the E–W azimuth suggests that instrumental

errors are relatively insignificant. For low P numbers caused by instrumental aberrations may be expected to show a rather uniform angular distribution. This is not the case in the present data, as shown by Fig. 12. Of particular importance are the 80°–90° sections, where the "accidental" counting rate of the telescope should contribute the greatest relative intensity. Of thirty-two sections for this range, only one exhibits a $P<0.05$; none a $P\lesssim 0.01$. That instrumental errors are unimportant is further indicated by the absence in the north–south data of many low P numbers.

We return now to the discussion of the differences between the N–S and E–W observations. It is important to know if this difference is a real cosmic-ray effect or if it can be accounted for by limitations in the experimental or analytical procedure.[25] These possibilities must be considered: (1) That the total time interval covered by the N–S data did not include magnetic or barometric changes like those which occurred while the E–W observations were made. (2) That, in view of the smaller number of curves for the N–S azimuth, there should be effectively a fractional increase in the number of constraints[26] pertaining to these data in the application of Pearson's test; for an implicit relationship between the reference (normalized time-average) curve and any individual short-time curve exists, since the reference curve is an average of all of the individual curves. (3) That a real difference in

[25] A re-examination of the data by another statistical method is desirable, but up to the date of this writing there has been no time for this. In a forthcoming paper by Schremp and Yeater, however, reference will be made to the results of this second analysis, which will have been carried out by then.

[26] The number of constraints for a section in either azimuth has been taken as one; this constraint enters because the total number of counts in the reference section is made equal to the total number in the section being tested for a time variation. "Constraint," in the sense implied here, is defined in Fry, reference 18.

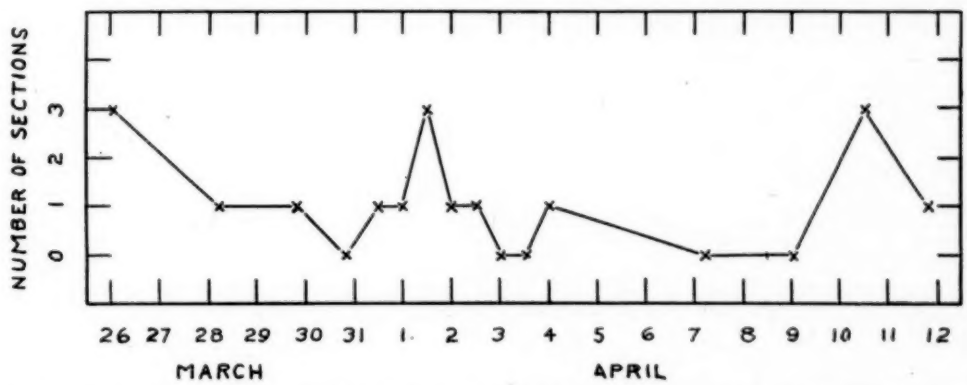

FIG. 14. Time distribution of the 10° sections, of individual E–W curves, for which $P\lesssim 0.01$. The time of day indicated for each date is noon.

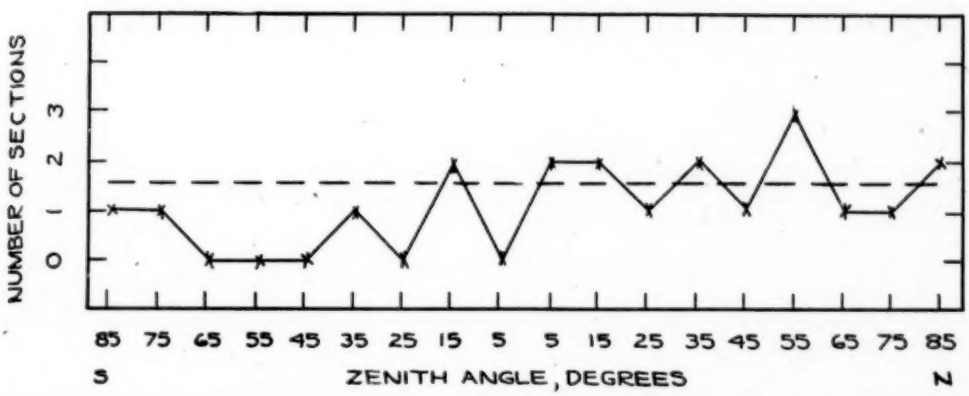

FIG. 15. Number of 10° sections of individual N–S directional intensity curves for which $P\lesssim 0.2$, as a function of zenith angle. The broken horizontal "chance line" shows the most probable number expected by chance for any 10° interval.

the nature of the time variations in the cosmic-ray directional intensity distribution exists between the N–S and E–W azimuths. All of these possibilities will now be examined.

(1) The complete time interval for the eight N–S curves includes five days as against eighteen days, March 25 to April 12, 1943, for the sixteen E–W curves. During these five days, from April 15 to April 20, 1943, the international magnetic character numbers (cf. Fig. 17) show no exceptional uniformity; no general difference between the magnetic fluctuations associated with the N–S and E–W cosmic-ray data is revealed. There is the particular difference, however, that no magnetic disturbances severe enough to be classed as storms occurred between April 15 and April 20; while three brief storms between March 23 and April 12 were reported by some of the stations making magnetic observations.[27] As to the atmospheric pressure, the barometer readings at St. Louis for the period April 15 to April 20 show a somewhat more stable condition than existed at several times during the preceding month (cf. Fig. 17). A change of pressure amounting to 0.6 cm Hg did occur, however, while the N–S cosmic-ray data were taken; the changes were not more than 1.0 cm throughout most of the time for the E–W data.

The possibility exists, therefore, that the N–S results may be caused by the somewhat more quiescent conditions of the magnetic field and/or atmospheric pressure; but the evidence favoring this possibility is not entirely convincing. For the

[27] Terr. Mag. 48, 117, 188 (1943).

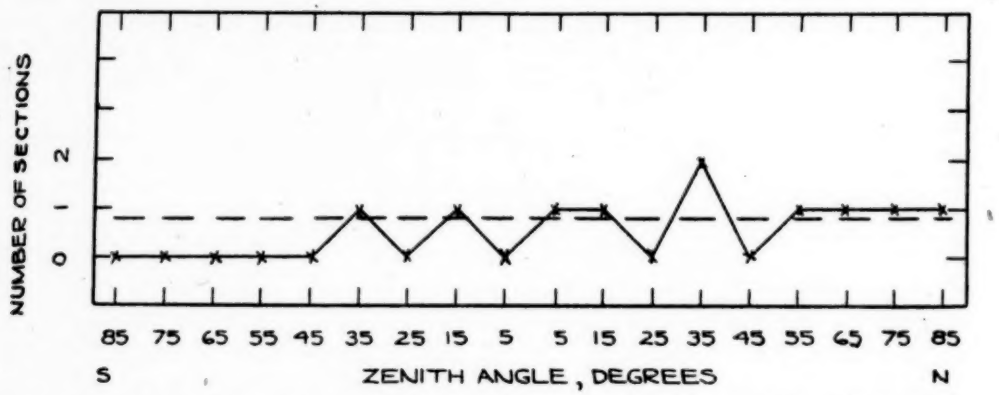

FIG. 16. Number of 10° sections of individual N–S directional intensity curves for which $P \leq 0.1$, as a function of zenith angle. The broken line shows the most probable number expected by chance for any 10° interval.

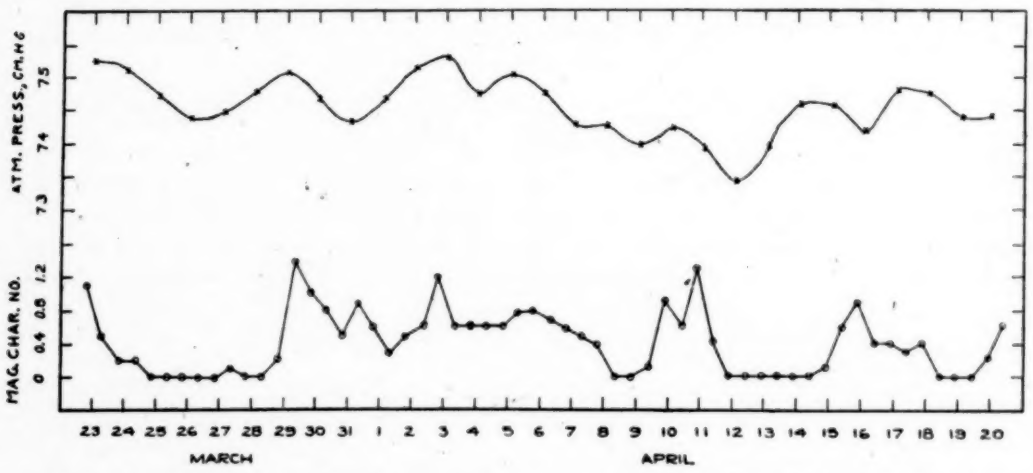

FIG. 17. Variation of atmospheric pressure at St. Louis and of international magnetic character number, for the time covered by the present cosmic-ray data. The time of day indicated for each date is noon. The scale of ordinates for each curve is given at the left.

difference between the N–S and E–W cases, with regard to magnetic and barometric conditions, is not great.

(2) Even if we take the somewhat extreme steps of increasing the number of constraints for the N–S case only, and letting this increase equal one, the difference between the N–S and E–W results is not removed. In this case an entry of $P \lesssim 0.1$ in Table II becomes $P \lesssim 0.05$; Fig. 16 then becomes a graph of $P \lesssim 0.05$, with the height of the "chance line" reduced from 0.8 to 0.4. For this assumed condition the average ratio of chance expectation to observed value for the number of sections with $P \lesssim 0.05$ is 0.72, and also for $P \lesssim 0.01$ it is 0.72, with no entries for $P \ll 0.01$; these are to be compared to values of 0.3 and 0.17, respectively, for the corresponding ratios in the E–W case, where in addition P values lower than 0.0001 are found (Table I).

From this it appears that the difference in the present results for the N–S and E–W azimuths has its origin in some physical circumstance.

(3) This possibility, that the character of the time variations in the cosmic-ray directional intensity patterns is a function of azimuth, is not ruled out by the preceding considerations (1) and (2). That time variations in the N–S pattern may be absent is consistent, furthermore, with certain interpretations of the E–W data, as discussed in the next section. This third possibility, in the light of all the evidence, appears by far the strongest.

3. Implications of the Time Variations

One of the purposes of this experiment was to determine whether rapid time variations exist for the anomalies observed in the static cosmic-ray directional intensity pattern. The result of the experiment in this respect is that no significant variations in these anomalies are revealed by the present analysis. On the other hand, this analysis does indicate time variations of an independent character.

The immediate evidence for the independence of the static and dynamic anomalies lies in the fact that the circular symmetry exhibited by the static anomalies is not exhibited by the time variations in the pattern. The fundamental reason for this independence, however, would appear to be that different origins can be ascribed to the two effects. The circularly symmetrical anomalies in the time-average curves have their origin in atmospheric effects, as previously discussed by others.[28] The probable origin of the time variations will be discussed in the following paragraphs.

In the E–W azimuth a dynamic character in the cosmic-ray directional intensity pattern is indicated by Fig. 12 to exist near W 65°, and E 75°, and near W 15° and E 25°. As we shall see, there is reason to believe that these peaks actually represent symmetrical dynamic regions. For if the variations in the pattern which produce the peaks in Fig. 12 really occur at 20° and 70°, respectively, then the observed asymmetry can easily result from a slight inaccuracy in the determination of the vertical direction for the cosmic ray telescope. An error of even a degree or two in this determination[29] can cause a time variation at 20° to appear at 25° on one side and 15° on the other in Fig. 12, and similarly for variations at 70°; for a point at E 25°, for example, means only that the variation in the pattern lies between 20° and 30°. Such an explaining away of the small observed asymmetry of course requires that this asymmetry be in the

TABLE II. Probability values P describing the goodness of fit of the individual north–south curves with their time average.

	Curve / Angular section	I	II	III	IV	V	VI	VII	VHI
S	90–80	0.37	0.95	0.86	0.18	0.24	0.62	0.66	0.64
	80–70	.64	.32	.66	.14		.50	.50	.85
	70–60	.48	.95	.57	.70	.77	.95	.92	.69
	60–50	.24	.87	.51	.60	.66	.52	.46	.84
	50–40	.91	.60	.33	.65	.89	.27	.24	.60
	40–30	.52	.32	.76	.08	.68	.58	.68	.36
	30–20	.25	.90	.29	.70	.28	.56	.22	.32
	20–10	.85	.52	.84	.17	.74	.52	.49	.06
	10–0	.98	.65	.93	.87	.28	.62	.84	.95
N	0–10	.46	.09	.32	.96	.19	.30	.82	.29
	10–20	.08	.20	.57	.69	.57	.74	.84	.58
	20–30	.64	.18	.62	.41	.74	.26	.56	.97
	30–40	.64	.26	.66	.10	.25	.09		.75
	40–50	.41	.86	.47	.71	.71	.95	.16	.46
	50–60	.41	.29	.17	.58	.027	.45	.49	.18
	60–70	.56	.21	.80	.90	.63	.80	.71	.07
	70–80	.72	.62	.30	.97	.02	.82	.33	.52
	80–90	.91	.68	.08	.18	.69	.88	.92	.56

[28] Cf. references 4, 6, and 7.

[29] The geomagnetic vertical should be used when magnetic effects are discussed, and this is about 3½° to the east of the geographic vertical at St. Louis; the geographic vertical, however, is used for the present data. If the geomagnetic vertical is used, the peaks in Fig. 11 tend to become symmetrical. The experimental alignment of the telescope with the vertical direction was accurate to about 1°.

same direction for both the 20° and 70° cases; this requirement is fulfilled.

Another feature to be noted in Fig. 12 is that the higher peaks occur at the larger angles. The greater tendency to change thus exhibited at the large zenith angle of 70° is an indication that magnetic effects, rather than atmospheric pressure changes, are responsible for these time variations. This follows because variations caused by atmospheric pressure changes should be more pronounced at smaller angles.[30] Variations caused by magnetic changes, as brought out later, are not restricted in this way.

Further evidence that these cosmic-ray time variations are associated with terrestrial magnetic changes comes from the time distribution of the very low values of the probability index P (Fig. 14). It happens that maxima in this curve, centered at March 26, April 1, and April 10, are observed coincident with or just following the occurrence of magnetic disturbances. (Magnetic data are plotted in Fig. 17; the disturbances referred to are listed specifically elsewhere.)[31]

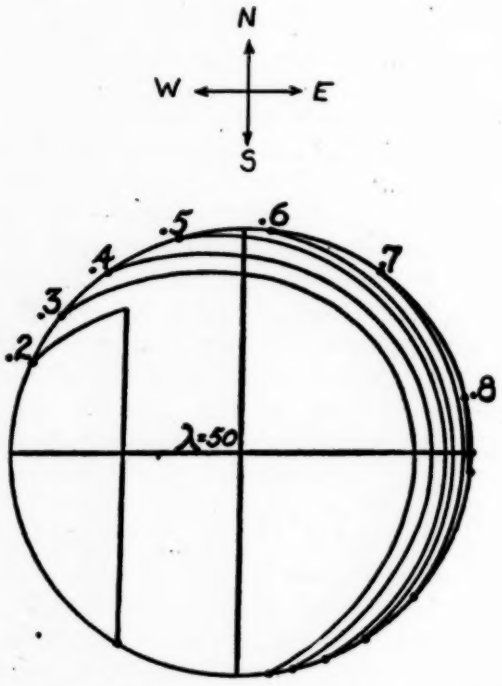

FIG. 18. Simple shadow cones, for positive particles, at geomagnetic latitude north 50°. Energies are in Störmers.

These magnetic changes can affect the directional pattern by distorting the allowed cones of the primary cosmic rays. In considering the effect of changes in these cones we perceive that the absence of time variations in the N–S pattern can be consistent with the hypothesis that variations at other azimuths are caused by magnetic fluctuations.

For directional cosmic-ray changes of magnetic origin must be caused by the movement of an allowed cone boundary across the direction of observation. Now at the magnetic latitude of St. Louis, approximately N 48°, the boundary of the allowed cone is essentially the common boundary of the simple shadow cone and the Störmer cone.[32] The character of this boundary for various energies is shown in Fig. 18.[33] From this we can see how certain of these boundaries can produce the observed dynamic regions in the E–W pattern (at approximately ±20° and ±70°) without intersecting the N–S meridian. At W 20°, for example, lies a cone boundary for a positive particle with an energy slightly greater than 0.2 Störmer. The N–S meridian is in the forbidden region for this energy; hence a movement of this cone can produce no changes in the N–S pattern. The region near E 20° can be similarly described with reference to a negative primary particle. In dealing with the region near E 70° we can invoke the cone for a positive particle with energy of about 0.6 Störmer. In this case the north–south meridian is in the region of full light, and no variations in the N–S pattern are to be expected from changes in this cone. The corresponding cone for a negative particle can account in a similar way for the dynamic character of the pattern near W 70°.

In thus relating the observed time variations with magnetic fluctuations, we have a reasonable explanation of their symmetry in the E–W azimuth and their absence in the N–S azimuth. More than this, we find in the symmetry of these variations a new kind of evidence supporting a conclusion previously advanced by Schremp,[34] to

[30] This can be demonstrated as follows: Let h_0 be the vertical height of the atmosphere, and h the atmospheric thickness looking in the direction θ from the zenith. Then to a good approximation

$$h = h_0 \sec\theta \quad \text{or} \quad \cos\theta = h_0/h.$$

Hence

$$(\partial\theta/\partial h_0)_{h=\text{const.}} = -(1/h)\ \text{cosec}\ \theta = -(1/h_0)\ \text{cotan}\ \theta.$$

[31] See reference 27.

[32] Cf. E. J. Schremp, Phys. Rev. **54**, 169 (1938).

[33] Reference 32, Fig. 5.

[34] Cf. reference 7.

the effect that positive and negative particles with the same e/m exist among the primary cosmic rays.

ACKNOWLEDGMENTS

Thanks are expressed especially to Dr. E. J. Schremp, who proposed the problem, for his cooperation during the experimental work and for his help in the preparation of this paper; a particular debt of gratitude is due him for suggesting the statistical analysis and for contributing to the interpretation of the results of this analysis. The encouragement and advice of Professor A. L. Hughes is gratefully acknowledged. The Department of Terrestrial Magnetism of the Carnegie Institution was helpful in supplying terrestrial magnetic data. Construction of the apparatus was financed jointly by Washington University and by grants made to Dr. E. J. Schremp from the Penrose Fund of the American Philosophical Society and from the Rumford Fund of the American Academy of Arts and Sciences.

APPENDIX I

Geiger Counter Construction

The cathode is shown in Fig. 19; this is entirely of brass. The end fittings marked *f* are used in the vacuum seal. The method of securing this seal is shown in detail in Fig. 19B; *R* is a gum rubber washer which is compressed against the inner surface of *f* and around the glass lead-in *g* to form the seal; the nut *N* bears on the metal washer *W* which crushes the rubber. The cathode is formed by hard-soldering the threaded fitting *f* to the disk *d*, then soft-soldering *d* to the light brass cylinder *c*. The glass lead-in is shown enlarged in Fig. 19C. The complete anode assembly is shown in Fig. 19D.

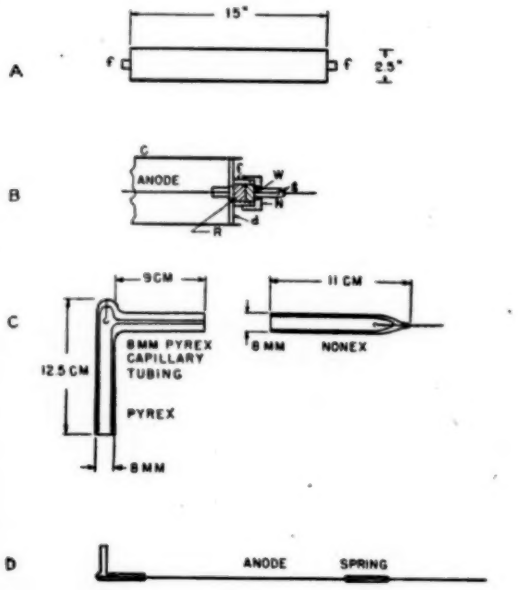

FIG. 19. Structure of the Geiger counters. A, cathode; B, vacuum seal; C, glass lead-ins; D, anode assembly.

The anode is prepared as follows: A length of 0.010" tungsten wire is cut off and a loop formed at each end by twisting under a tension of about two kilograms; the wire is then rubbed vigorously with fine sand paper. It was not found necessary to flash the anodes in vacuum.

The cathodes are cleaned in an acid solution consisting of about three parts HNO_3 (sp. gr. 1.403), one part H_2SO_4 (sp. gr. 1.84) and twenty parts water. After being washed several times in tap water and in distilled water the cathodes are permitted to stand in air a few days to form a thin oxide layer on the brass.

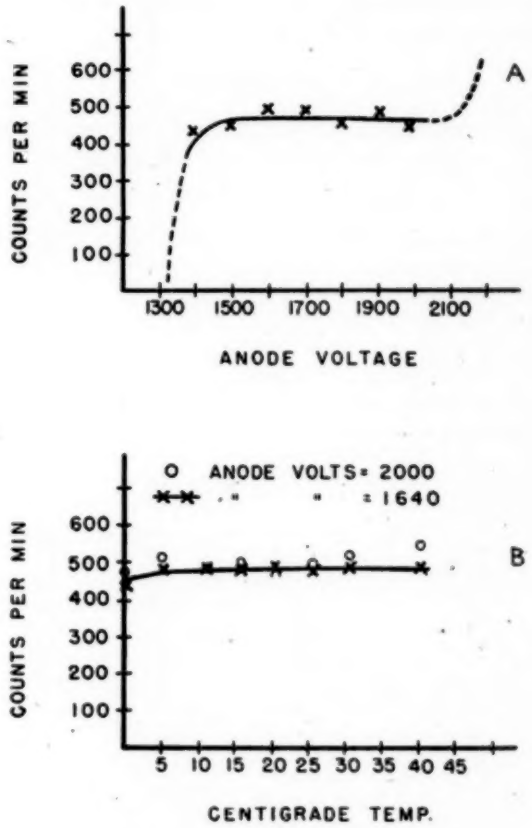

FIG. 20. A, Geiger counter characteristic curve; B, counting rate of Geiger counter at different temperatures.

The filling gas mixture consists of 8 mm alcohol, 4 mm petroleum ether, and 7.8 cm argon. A typical characteristic curve is shown in Fig. 20A. The effect of temperature on the counting rate of an early test counter is shown in Fig. 20B.

APPENDIX II

Pulse Cable Connection

The manner in which the counter pulses are properly distributed to the mixing circuits is indicated in Fig. 21 and Fig. 22. Figure 21 is a simplified diagram showing the interconnections for just two of the eighteen directions. From each quenching unit a shielded cable runs one-fourth of the way around the drum on either side, and from these cables the pulses are tapped off to the coincidence circuits. The mixing circuits for the direction indicated by the broken lines are represented by the ten small dots at the left, while for the direction defined by the full lines the mixing circuits are at the right. Counter units A and A' feed into vacuum tubes a_1 and a_1'; similarly B and B' feed into b_1 and b_1', etc. In order to interconnect thirty-six counter units for eighteen directions this system is simply extended; one cable from each counter unit runs halfway around the drum and pulses are tapped off at the appropriate mixing-recording panel. The complete wiring diagram for this is given in Fig. 22.

APPENDIX III

Suggested Improvements In Apparatus and Procedure for Cosmic-Ray Directional Intensity Measurements

(1) The general circuit arrangement of the multidirectional telescope is satisfactory, but a recording circuit not using neon lamps should be employed.

(2) The "fast" type of Geiger counter containing organic vapor is not good for this work because of limited life.

(3) The cosmic-ray telescope should be operated in a shack in which constant conditions of temperature and illumination can be maintained.

(4) An output pulse length, for the Geiger counter-quenching circuit combination, of the order of a microsecond should be achieved, in order to reduce the "accidental" counting rate. This, it is believed, can be done by adjustment of constants in the quenching circuit; a resulting excessive decrease in pulse size can be compensated by another amplifier stage. This seems more practicable than a triple coincidence counting system, for the present type of apparatus.

(5) Observations should be made in rapid alternation in the N–S and E–W azimuths (or any other pair at right angles to each other) so that in effect patterns for both azimuths are recorded simultaneously.

(6) If measurements like those described in this paper are made, a second telescope operating at fixed zenith angles should be run concurrently, giving direct readings of intensity *versus* time for various angles. If possible, ionization chamber measurements of total intensity should be carried along at the same time.

(7) A high angular resolution, of three degrees or less, with respect to zenith angle, is desirable for a cosmic-ray telescope used in tracing the fine structure; for a more precise knowledge of this pattern, which would immeasurably facilitate the interpretation of the data, appears to be easily attainable by this means. A second telescope used for direct observation of time variations at fixed angles would preferably have a resolution in zenith angle equal to the angular spacing between the directional trains, in the present case ten degrees.

APPENDIX IV

Notes on Control Circuit

These notes pertain to Fig. 2. In reference 13 is given a description of the manner of starting the motor M (which causes the drum to rotate) at optional time intervals of

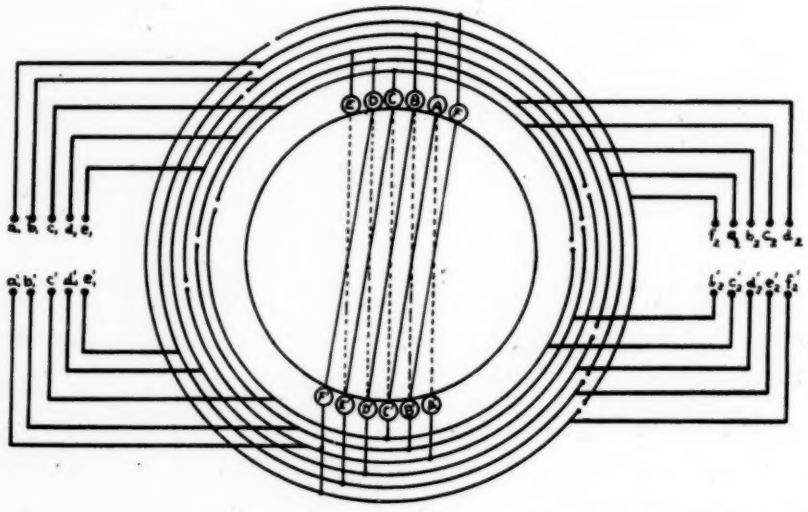

FIG. 21. Pulse cable connections for two directions of observation. The dots marked with lower case letters represent mixing tubes in the recording circuit. Circles marked with capital letters represent Geiger counter units. The angular separation of these two directions is 10°.

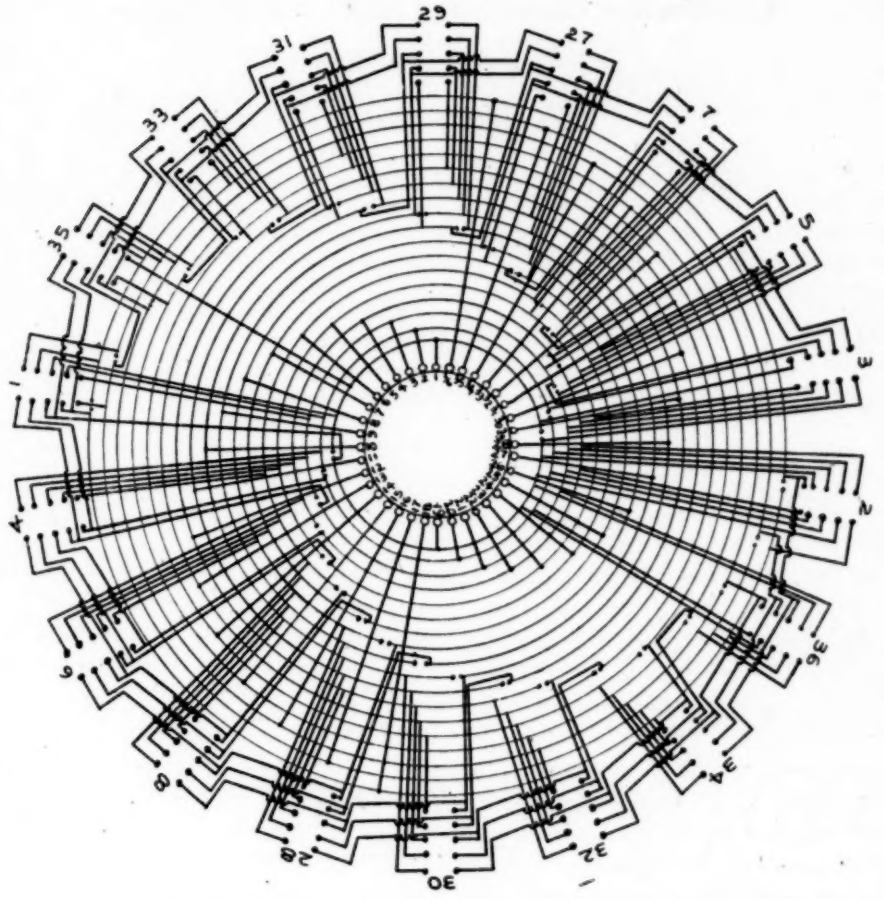

FIG. 22. Complete pulse cable connections for the multidirectional cosmic-ray telescope. The Geiger units are numbered *1* to *36* around the inner circle. The cables are shown as concentric semicircles. Every group of dots represents one panel of mixing tubes. Each group is assigned the number of one of the Geiger units of the center train for that direction; for example, group 1 represents the mixing-recording circuit for the complete directional array of counters which includes the following units in double-coincidence trains: *3* and *17*, *2* and *18*, *1* and *19*, *36* and *20*, *35* and *21*.

twenty minutes, thirty minutes, or one hour, and of the way in which the motor is stopped after it has experienced a predetermined number of revolutions. This number, and hence the angle through which the drum is rotated, is chosen by selector switch Sw_2. At the position 2 of Sw_2, relay 2 will operate, and hence the motor will stop, every time either pair of contacts *a* or *d* of Sw_3 is shorted by the rolling contact* which is driven by *M* through reducing gears. This gear ratio is so adjusted that for a six degree rotation of the drum the rolling contact passes once over every set of contacts *a* to *f*, and these sets are spaced sixty degrees apart. Hence for the setting 2 of Sw_2 the motor is stopped after every three-degree rotation of the drum. Similarly, for setting 1 the motor is stopped whenever any set of contacts *a* to *f* is shorted, or for every one degree rotation of the drum. Setting 3 of Sw_2 stops the rotation of the drum after six degrees.

When relay 2 and simultaneously relay 4 are activated, thereby stopping the motor, the coils *MR* and *L* are energized. *MR* is the coil of a message register which counts the number of small rotations of the drum and hence indicates the angular position of the drum at any time during a run of the experiment. *L* is a solenoid which trips a camera and causes a photographic record of the meter panel to be made.

The camera can be tripped automatically by the clock switch without any rotation of the drum. This occurs if Sw_4 is at position 2.

Switch Sw_1 is a push-button switch by means of which the clock switch can be shorted by the operator, thereby enabling him to start the motor *M* or, if Sw_4 is at 2, to take a picture of the meter panel, at any arbitrary time.

* Cf. reference 13.

PHYSICAL REVIEW VOLUME 67, NUMBERS 3 AND 4 FEBRUARY 1 AND 15, 1945

The Nuclear Excitation of Silver and Cadmium

MARCELLUS L. WIEDENBECK
Department of Physics, University of Notre Dame, Notre Dame, Indiana
(Received December 9, 1944)

The x-ray and electron excitation of silver and cadmium has been studied in the region from their thresholds to 3.3 Mev. The thick target x-ray excitation curve for silver shows the threshold to be 1.18±0.03 Mev and higher activation levels are found at 1.59, 1.95, 2.32, 2.76, and 3.13 Mev, respectively. The half-life of the metastable state of silver is 40.4±0.2 seconds. Both the thick target x-ray excitation curve and the electron excitation curve for cadmium indicate that the threshold for these processes is 1.25±0.03 Mev while other activation levels are found at 1.68, 2.08, and 2.56 Mev. The energy of the metastable level of cadmium, as determined from data on the absorption in aluminum of the conversion electrons emitted in the process, has been found to be 195 kev and the half-life of the metastable level is 48.7±0.3 minutes.

INTRODUCTION

NUCLEAR isomerism was first observed when it was found that three radioactive periods could be produced by the irradiation of bromine with slow neutrons.[1] Since there are only two stable isotopes of bromine, namely, $_{35}Br^{79}$ and $_{35}Br^{81}$, it was necessary to assign two of these periods to a single isotope.

It was pointed out by von Weizsacker[2] that such a condition may be explained by the assumption that there is a large difference of spin between the two states and hence the state of higher energy may exhibit a relatively long lifetime against γ-emission. In 1938, Goldhaber, Hill, and Szilard[3] found that the *stable* isotope In^{115} possessed such a metastable level which could be excited by fast neutrons and which decayed to the ground level with a period of 4.1 hours. Since this time, eleven stable nuclei possessing metastable states have been reported.[4]

The presence of a low lying metastable state in a stable nucleus affords a very accurate method for determining a series of the energy levels of that nucleus. Inasmuch as the transition from the metastable to the ground level is forbidden, because of the large difference in angular momentum, the direct transition from the ground state to the metastable state is just as strongly forbidden. Therefore, to reach the metastable state, the nucleus must first be raised to a higher activated level, the spin of which is intermediate between that of the ground and metastable states. The nucleus can then leave this activated level either by direct transition to the ground state or it may fall to the metastable level with a further change of spin, $l_m - l_a$, and the emission of energy, $h\nu_a - h\nu_m$. After the nucleus is in the metastable state it can only fall to the ground state with the emission of energy $h\nu_m$ and a large change of angular momentum, $l_g - l_m$.

Nuclei, in general, can be raised to the activated levels by inelastic collision with neutrons,[3] protons,[5] alpha-particles,[6] electrons,[7] and by the absorption of x-rays. The latter two processes afford more complete and accurate methods for studying the excitation and for the determining of the positions of the higher activation levels.

If x-rays are produced by a beam of monochromatic electrons striking a thin target, the intensity of a given isochromat, in the continuous x-ray spectrum thus produced, was shown, in the energy ranges concerned here,[8] to remain constant at accelerating voltages equal to or greater than the energy of this isochromat. Hence, one should expect the activity in a sample irradiated with thin target x-rays to rise quickly to a certain value when the threshold energy for the process is reached and then to remain essentially constant

[1] C. H. Johnson and F. T. Hamblin, Nature **138**, 504 (1936); A. H. Snell, Phys. Rev. **52**, 1007 (1937).
[2] C. F. von Weizsacker, Naturwiss. **24**, 813 (1936).
[3] M. Goldhaber, R. D. Hill, and L. Szilard, Phys. Rev. **55**, 47 (1939).
[4] For a recent reference list see: G. T. Seaborg, Rev. Mod. Phys. **16**, 1 (1944)
[5] S. W. Barnes and P. W. Aradine, Phys. Rev. **55**, 50 (1939).
[6] K. Lark-Horovitz, J. R. Risser, and R. N. Smith, Phys. Rev. **55**, 878 (1939).
[7] G. B. Collins and B. Waldman, Phys. Rev. **57**, 1088 (1940).
[8] E. Guth, Phys. Rev. **59**, 325 (1941).

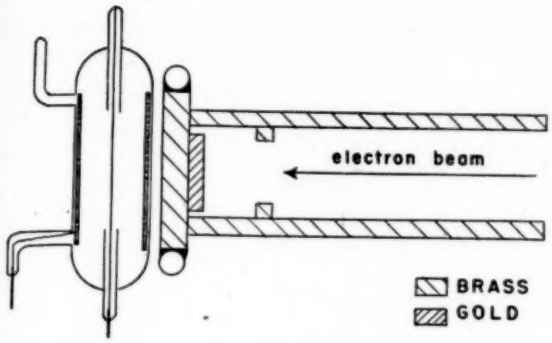

FIG. 1. Thick x-ray target arrangement with silver counter in bombardment position.

until a second activation level is reached, whereupon the activity will again rise in a similar manner. The positions of these sudden rises in the activity *vs.* accelerating potential curves then indicate the energy of the activated levels.

Similarly, as has been pointed out by Guth,[8] if we consider the thick target continuous x-ray spectrum as composed of a multitude of thin targets superimposed one behind the other, the activity should rise linearly with accelerating potential until another activation level is reached, whereupon the slope of the activity curve increases. The intersection of the two linear portions thus indicates the position of the activation level.

If the process is one of resonance absorption, the electron excitation curves should show sharp peaks when an activation level is reached and then fall to a lower value once this energy has been exceeded, similarly to the well-known excitation curves of Franck and Hertz for electrons striking an atom.

EXPERIMENTAL PROCEDURE

A. Silver

It had previously been found[9] that silver possesses a metastable state 93 kev above the ground level which decays with a half-life of 40 seconds. It has also been shown that this activity can be

[9] L. W. Alvarez, A. C. Helmholz, and E. Nelson, Phys. Rev. **57**, 660 (1940); G. E. Valley and R. L. McCreary, Phys. Rev. **56**, 863 (1939); A. C. Helmholz, Phys. Rev. **60**, 160 (1941); E. Segrè and G. T. Seaborg, Phys. Rev. **59**, 212 (1941).

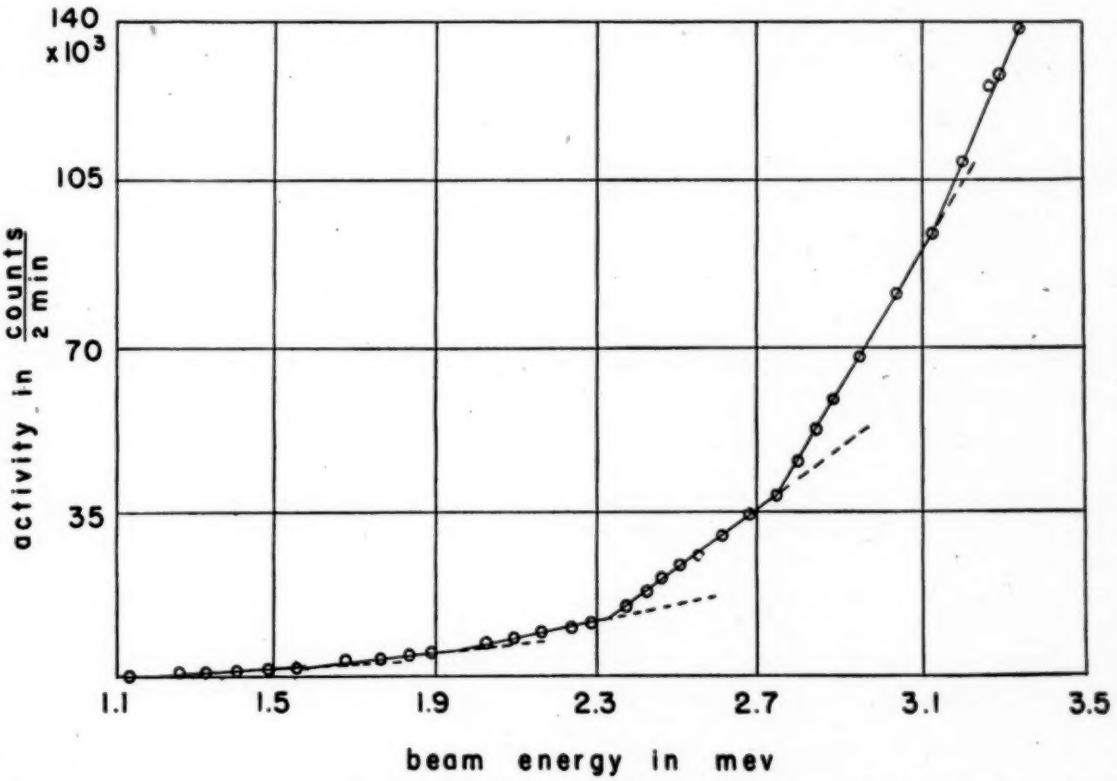

FIG. 2. Thick target x-ray excitation curve for silver. The positions of sudden changes in slope give the position of the energy levels which combine with the metastable state.

produced by direct x-ray bombardment of silver.[10]

In the present work, the x-rays were produced by a beam of electrons impinging on a gold target. The electrons were accelerated by a high pressure van de Graaff-Herb generator.[11] The sample of silver being irradiated formed the

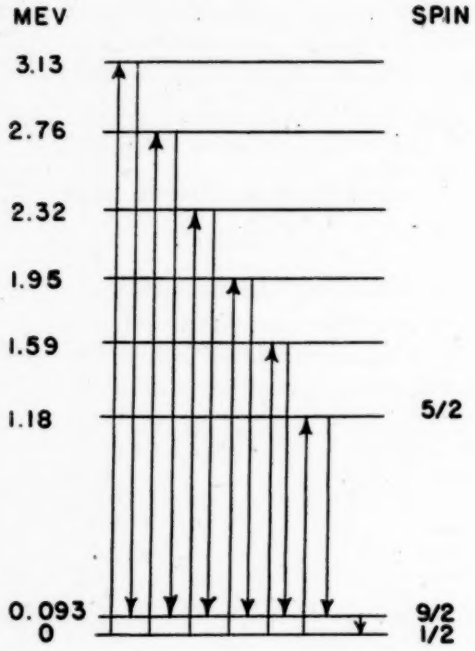

FIG. 3. Nuclear energy level diagram for silver showing the various transitions by which nuclei may reach the metastable state.

cylinder of a counter which was placed directly in front of the target as shown in Fig. 1.

The counters used were self-quenching and were filled with an argon (90 percent)—ether (10 percent) mixture. It has been found that such a counter could be irradiated for several hours with very intense x-ray beams without showing a change in threshold or counting efficiency. Argon-alcohol counters, on the other hand, quickly deteriorate under the action of an intense x-ray beam.

In obtaining an excitation curve it is necessary to find the activity produced as a function of the accelerating voltage, keeping all other quantities which might influence the activity constant. For the work with thick target x-rays, the electron beam used was 200 microamperes, the irradiation time was fixed at two minutes, and the activity was taken as the number of counts above background during the two-minute period starting five seconds after the beam was turned off.

The curve thus obtained is shown in Fig. 2. It can be seen that the threshold, which represents the first activation level above the metastable state, is found at 1.18 Mev and higher activation levels, indicated by the sharp breaks in the curve, are found at 1.59, 1.95, 2.32, 2.76, and 3.13 Mev, respectively.

From these data it is possible to draw a nuclear energy level diagram for silver (Fig. 3) indicating the excitation and the subsequent decay of the excited levels combining with the metastable state.

Difficulties were encountered in observing the excitation of silver by direct electron bombardment since the energy of the conversion electrons emitted in the transition from the metastable to the ground state is very low (66 kev). These electrons are completely stopped by a very small thickness of material. As a consequence, it was necessary to find an arrangement by which the silver could be bombarded and then placed directly inside of a counter with as short a time delay as possible. To accomplish this, a counter was mounted on a large brass stopcock (Fig. 4) along with a ballast bottle filled with an argon-ether mixture at the same pressure as that in the counter. With this arrangement, the sample was placed in a cavity (*a*) in the inner part of the stopcock when at position (*b*). After the cavity was evacuated at position (*c*) by means of a fore-pump, it was turned to position (*d*). The cavity was then filled with the gas mixture to the same

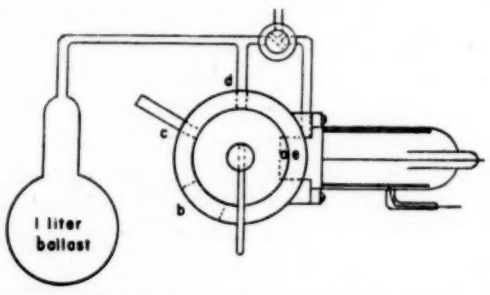

FIG. 4. Geiger-Müller counter mounted on a brass stopcock for admitting active sample to the inside of the counter.

[10] J. Feldmeier and G. B. Collins, Phys. Rev. 59, 937 (1941).

[11] This name seems appropriate for generators of the high pressure type since much of the early work in their development was done by Herb and his group.

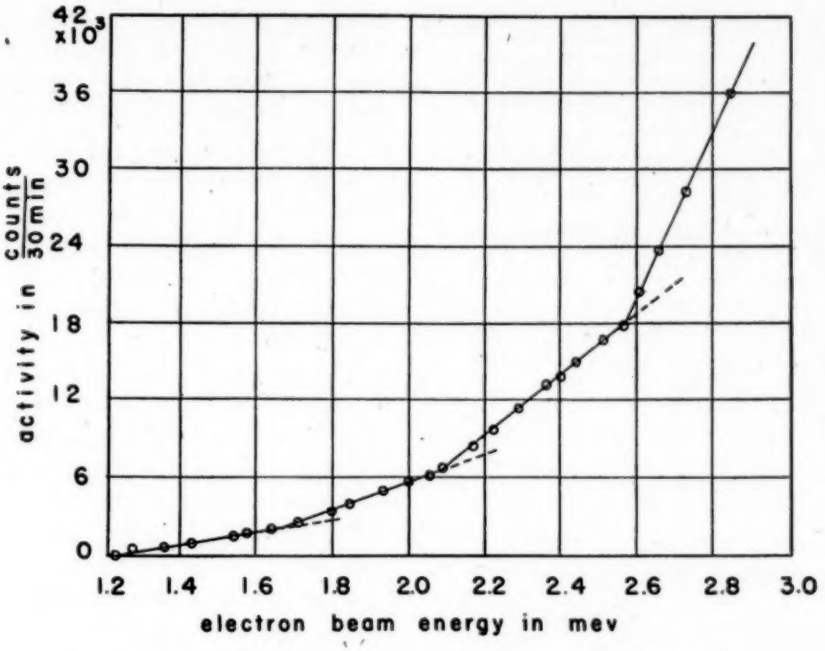

FIG. 5. Thick target x-ray excitation curve for cadmium.

pressure as the counter after which it was turned to position (*e*) within the counter. With this arrangement the sample could be gotten into the counter within two minutes after irradiation was stopped. Activity produced by the electron bombardment was observed but the complete excitation curve has not as yet been studied for this element.

The half-life of the metastable state was determined from a decay curve extending through six half-life periods with an initial activity of 8000 counts/15 seconds. These data give a value of 40.4 ± 0.2 seconds for the half-life.

This metastable state may belong either to $_{47}Ag^{107}$ or $_{47}Ag^{109}$ both of which have spins of $\frac{1}{2}$ in the ground state. From the decay period and energy of the metastable state, this would imply a spin of 9/2 for the metastable state[12] and a probable spin of 5/2 for the level at 1.18 Mev.

[12] A. C. Helmholz, Phys. Rev. 60, 415 (1941).

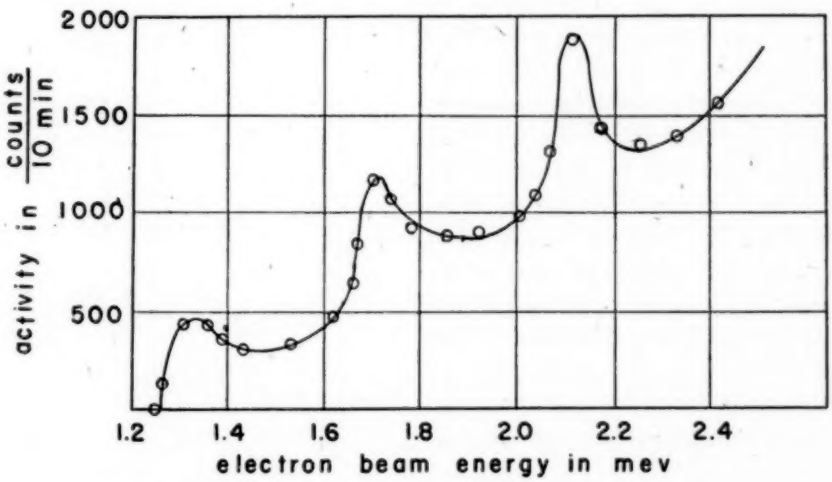

FIG. 6. Electron excitation curve for cadmium. The positions of sharp rises give the energy of the excited levels in the resonance absorption.

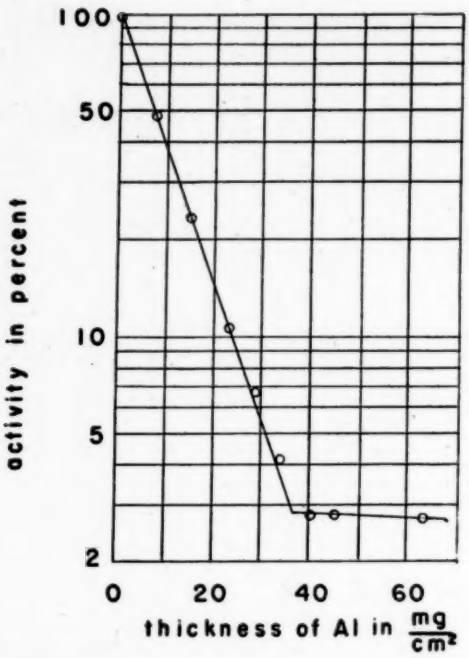

FIG. 7. Absorption curve for the conversion electrons emitted in the transition: $Cd^* \rightarrow Cd + \gamma$.

B. Cadmium

The fifty-minute period of cadmium[10,13] has been studied in a manner similar to that used with silver. The positions of the activated levels have been found from a thick target x-ray excitation curve. The activity was taken as the number of counts obtained during the first half-hour after irradiating a counter with a cadmium cathode for a time of ten minutes with a beam current of 250 microamperes. This curve is shown in Fig. 5 and as is seen, the threshold for this reaction, $Cd + \gamma \rightarrow Cd^*$, is 1.25 ± 0.03 Mev and higher activation levels are found at 1.68, 2.08, and 2.56 Mev, respectively. In the region of the threshold, the samples were irradiated for much longer times with beam currents up to 700 microamperes so that the activity of the cadmium would greatly exceed the normal background of the counter. These activities were then corrected to the proper value of irradiation time and beam current.

An electron excitation curve for cadmium was also obtained by bombarding thin samples of cadmium directly with the electron beam. The activity produced in the samples was measured by the counter previously described (Fig. 4). The excitation curve (Fig. 6) shows a series of rather sharp peaks superimposed on a continuous background. The positions of these peaks occur at the same voltages as the breaks in the thick target x-ray excitation curves. The relative sharpness of these peaks in the electron excitation curve gives further evidence that the process is one of line absorption.

The energy of the metastable state was also determined by measuring the absorption of the conversion electrons in aluminum foils. Several runs were taken at a constant accelerating voltage and with constant beam current with various thicknesses of aluminum between the cadmium and the counting region. The activity was thus found as a function of the thickness (Fig. 7). The energy of the metastable state is then the energy of the conversion electrons plus the binding energy of the K electrons of cadmium. The value thus found is 195 kev.

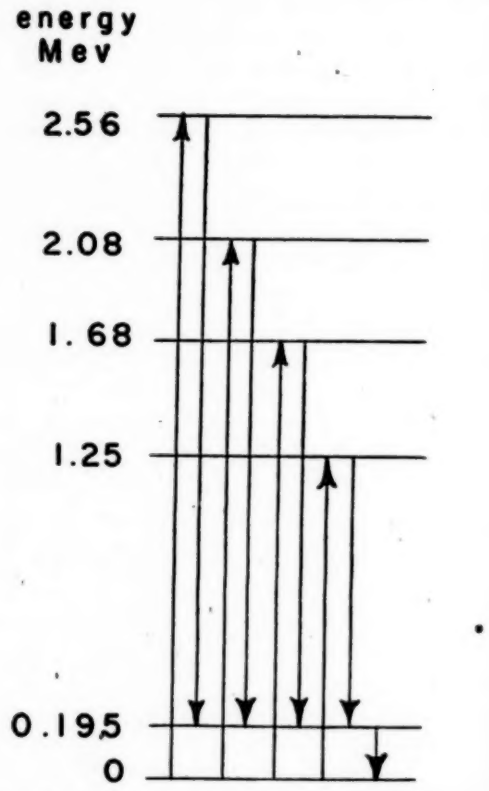

FIG. 8. Nuclear energy level scheme for cadmium.

[13] M. Dode and B. Pontecorvo, Comptes rendus 207, 287 (1938); M. L. Wiedenbeck, Bull. Am. Phys. Soc. 19, No. 3 (1944).

From these data one can draw a nuclear energy level diagram for cadmium (Fig. 8) showing the various energy levels combining with the metastable state.

The isotope of cadmium which possesses the metastable state can be either $Cd^{110,111,112,113}$ since this activity has also been produced by Scherrer *et al.*[14] by the Cd $\gamma-n$ reaction. Of these, Cd^{110} can probably be eliminated in as much as material with this period cannot be grown from silver having the 22-second period.

DISCUSSION

It is of interest to notice that metastable levels were found for two "triads:" silver, cadmium, and indium;[15] and platinum,[16] gold,[17] and mercury.[18] It is very likely that the group, krypton,[19] strontium,[20] and rubidium, may form another such "triad."

The level schemes in the "triad," silver, cadmium, and indium, show that the spacing of these levels is approximately uniform in any one of the elements. A similar condition has been found to exist in the case of nitrogen.[21] However, it should be noted that nitrogen is among the light elements while the elements studied here are of intermediate weight. It is possible that with the extension of these excitation curves to higher energies the levels will converge as in the case of atomic spectra.

ACKNOWLEDGMENTS

The author wishes to express his thanks to Mr. V. C. Stock and Brother Camille, C.S.C., of the Physics Department Machine Shop for their aid in the construction of apparatus. It is also with pleasure that the author expresses his thanks and appreciation to Dr. Eugene Guth for his many helpful suggestions in the analysis of this work.

[14] O. Huber, O. Lienhard, P. Scherrer, H. Wäffler, Helv. Phys. Acta **16**, 228 (1943).

[15] B. Waldman and M. L. Wiedenbeck, Bull. Am. Phys. Soc. **17**, No. 5 (Nov. 1942); M. L. Wiedenbeck, Thesis on nuclear excitation of Indium. Both thick and thin target x-ray excitation curves and the electron excitation curve has been obtained for indium and will be reported in the near future.

[16] R. Sherr, K. T. Bainbridge, and H. H. Anderson, Phys. Rev. **60**, 473 (1941).

[17] Data on gold will be published in a forthcoming issue of Phys. Rev.

[18] F. A. Heyn, Nature **139**, 842 (1937); E. McMillan, M. Kamen, and S. Ruben, Phys. Rev. **52**, 375 (1937); M. L. Pool, J. M. Cork, and R. L. Thornton, Phys. Rev. **52**, 239 (1937).

[19] A. Langsdorf, Jr., and E. Segrè, Phys. Rev. **57**, 105 (1940).

[20] L. A. Dubridge and J. Marshall, Phys. Rev. **56**, 706 (1939).

[21] P. Comparat, Nature **153**, 720 (1944).

PHYSICAL REVIEW VOLUME 67, NUMBERS 3 AND 4 FEBRUARY 1 AND 15, 1945

Approximate Partition Function in Generalized Bethe's Theory of Superlattices

J. S. WANG
Department of Physics, National Tsing Hua University, China*
(Received September 20, 1944)

An approximate expression for the partition function in a generalization of Bethe's theory of superlattices including long range interaction is obtained as an integral of the approximate energy expression in the generalized Bethe's theory. An alternative form of the energy expression is also considered, and higher approximations are treated. The paper ends with an application to the problem of adsorption.

1. INTRODUCTION

SINCE the appearance of Bragg and Williams' first paper many methods for the approximate solution of the problem of superlattices have been devised,[1] of which the method of Bethe is quite a good one. In Bethe's method the equilibrium values of the superlattice order and other quantities are obtained by an indirect method without evaluating the partition function. In some problems, such as the critical condensation phenomenon in adsorption, the true critical temperature in a face-centered cubic lattice, and the separation into phases studied by Easthope,[2] the value of the partition function itself is indispensable for their investigation, and Bethe's theory becomes powerless here.

Some time ago the author put forward a theory[3] of obtaining an approximate expression for the partition function in Bethe's theory, in its generalized form including long range interactions. Upon closer investigation it is found that the theory has two defects. First, the theory assumes that Bethe's method is consistent when applied to the case of neighbor interaction so that the condition of integrability is satisfied and a partition function always exists. Although this is true in the cases considered then, it is found that in the case of Bethe's second approximation and in the case of face-centered cubic lattice this is not true. Secondly, the numerical work involved in the calculation of the partition function is too laborious to be of any value in practical application.

In this paper a modification of the theory is given. The present theory is perfectly general and applicable in all cases. Moreover, the numerical work has been lessened considerably so that practical applications are possible. The general theory is given in Section 3 and Section 4 below.

In the application of Bethe's theory one works with what may be called the local grand partition function, instead of the ordinary partition function. The use of the grand partition function in the theory of superlattices has been justified in a general way in the paper of the author cited above. It is found, however, that a slight modification of the theory there given helps to clarify some points concerned with the procedure of obtaining the approximate partition function from Bethe's theory. This is done in Section 2.

In the problem of adsorption the number of adsorbed atoms depends on the presence of the gas from which the adsorbed atoms come, and it is evident that the grand partition function should be used. But since the grand partition function is also used in the theory of superlattices, the theory of adsorption is mathematically the same as the theory of superlattices. The corresponding quantities in the two theories have, however, different meanings, and some of them have to be treated differently. Thus the problem of adsorption receives a special treatment in Section 5.

The notation used in the present paper is nearly the same as that employed in the previous one. To facilitate the reading of the paper it is perhaps better to explain some of the basic quantities first. We consider a binary alloy composed of two kinds of atoms A and B with a total

* At present a member of National Southwest Associated University, Kunming, China.

[1] See a general review by F. C. Nix and W. Shockley, Rev. Mod. Phys. **10**, 1 (1938).

[2] C. E. Easthope, Proc. Camb. Phil. Soc. **34**, 68 (1938).

[3] J. S. Wang, "Free energy in the statistical theory of order-disorder transformation," Science Report of National Tsing Hua University [A], **IV**, Nos. 4–6 (April, 1941), (printed but failed to appear).

number N. The number of A atoms will be denoted by $N\theta$, and that of B atoms by $N(1-\theta)$. The N sites occupied by the N atoms are divided into two classes α and β, the number of α-sites is Nr and that of β-sites $N(1-r)$. The value of r is $\frac{1}{2}$ for a simple cubic and a body-centered cubic lattice, and is $\frac{1}{4}$ for a face-centered cubic lattice ($r=\frac{1}{2}$ is also possible, e.g., AuCu). To simplify the writing we sometimes use the abbreviation N_α for Nr, and N_β for $N(1-r)$. In any given state of order let there be $Nr\theta_\alpha$ A atoms in α-sites and $N(1-r)\theta_\beta$ A atoms in β-sites. Since the total number of A atoms is $N\theta$, we have

$$r\theta_\alpha+(1-r)\theta_\beta=\theta. \tag{1}$$

The degree of order s is defined by

$$s=\theta_\alpha-\theta_\beta. \tag{2}$$

This definition of s reduces to that of Bragg and Williams in the special case of $\theta=r$.

Other symbols will be explained when they are first introduced.

Recently a very interesting generalization of the quasi-chemical method of Fowler and Guggenheim has been made by Mr. C. N. Yang.[4] This method is very powerful, and perhaps Bethe's method will be superseded by it. The chief merit of the method is that the approximate partition function can be easily obtained as a closed expression. Mr. Yang is now making further investigation of the method.

2. THE GRAND PARTITION FUNCTION

The (configuration) energy of the alloy in the general case of long range interaction has been found to be[5]

$$\begin{aligned} &E=\tfrac{1}{2}\sum Z_\nu V_\nu^{BB}+\theta\sum Z_\nu(V_\nu^{AB}-V_\nu^{BB})+E_1,\\ &E_1=\sum M_\nu^{AA}V_\nu, \quad V_\nu=V_\nu^{AA}+V_\nu^{BB}-2V_\nu^{AB}, \end{aligned} \tag{3}$$

where M_ν^{AA} is the number of AA pairs of atoms situated at a distance a_ν apart, $\nu=1, 2, 3, \cdots$; V_ν^{AA}, V_ν^{BB}, V_ν^{AB} are the interaction energies of AA, BB, AB pairs at the distance a_ν; Z_ν is the sum of z_ν for all sites of the crystal, z_ν being the number of sites situated at a distance a_ν from a given site. The sign $\sum$ in (3) denotes a summation over the values of ν.

From the above expression for the energy the ordinary partition function is obtained as

$$\begin{aligned} P_\theta=&\exp\left(-\tfrac{1}{2}\sum Z_\nu V_\nu^{BB}/kT\right)\\ &\times\{\exp\left[-\sum Z_\nu(V_\nu^{AB}-V_\nu^{BB})/kT\right]\}^\theta\\ &\times\sum Q(\theta_\alpha,\theta_\beta,T), \end{aligned} \tag{4}$$

$$Q(\theta_\alpha,\theta_\beta,T)=\sum\gamma_M\exp\left(-\sum M_\nu^{AA}V_\nu/kT\right),$$

where γ_M is the number of arrangements for given values of θ_α, θ_β, M_ν^{AA}. The summation of $Q(\theta_\alpha,\theta_\beta,T)$ in the expression for P_θ is taken over all θ_α and θ_β satisfying the relation (1) with a fixed θ, and the summation in the expression for $Q(\theta_\alpha,\theta_\beta,T)$ is taken over all possible values of M_ν^{AA} for given θ_α and θ_β.

The purpose of the present paper is to obtain an approximate expression for $Q(\theta_\alpha,\theta_\beta,T)$ in Bethe's method.

The equilibrium values of θ_α, θ_β, and E_1 are given in terms of Q by the usual formulae

$$\frac{\partial}{\partial s}\log Q=0,$$

or

$$\frac{1}{r}\frac{\partial}{\partial\theta_\alpha}\log Q=\frac{1}{1-r}\frac{\partial}{\partial\theta_\beta}\log Q, \tag{5}$$

$$E_1=kT^2\frac{\partial}{\partial T}\log Q. \tag{6}$$

These equations can be deduced from the standard formula in terms of P_θ on replacing P_θ by its maximum term Q.

Let us now form the grand partition function K by considering the number of A atoms as variable. We have

$$\begin{aligned} &K=\sum\lambda^{N\theta}P_\theta=\exp\left(-\tfrac{1}{2}\sum Z_\nu V_\nu^{BB}/kT\right)K_1,\\ &K_1=\sum\xi^{N_\alpha\theta_\alpha+N_\beta\theta_\beta}Q(\theta_\alpha,\theta_\beta,T),\\ &\xi=\lambda\exp\left[-\sum Z_\nu(V_\nu^{AB}-V_\nu^{BB})/NkT\right], \end{aligned} \tag{7}$$

where the summation in K_1 is taken over all possible values of θ_α and θ_β. In order to obtain the equilibrium values of θ_α, θ_β, E_1 from the grand partition function, it is more convenient to introduce two ξ's, ξ_α and ξ_β, say, such that K_1 becomes

$$K_1=\sum\xi_\alpha^{N_\alpha\theta_\alpha}\xi_\beta^{N_\beta\theta_\beta}Q(\theta_\alpha,\theta_\beta,T). \tag{8}$$

[4] C. N. Yang, Chinese J. Phys. (in press).

[5] Equation (3) is given in J. S. Wang, reference 3, Eq. (36). It can be easily obtained by the procedure given in J. S. Wang, Proc. Roy. Soc. **A168**, 56 (1938), especially p. 58.

Then the usual formulae for the equilibrium values of θ_α, θ_β, E_1 are

$$Nr\theta_\alpha=\xi_\alpha\frac{\partial}{\partial\xi_\alpha}\log K_1,$$
$$N(1-r)\theta_\beta=\xi_\beta\frac{\partial}{\partial\xi_\beta}\log K_1, \tag{9}$$

$$E_1=kT^2\frac{\partial}{\partial T}\log K_1, \tag{10}$$

and in the final result we must put

$$\xi_\alpha=\xi_\beta=\xi. \tag{11}$$

To prove that the equilibrium values determined from the grand partition function K_1 by means of (9)–(11) are the same as those calculated from the ordinary partition function Q by means of (5) and (6), we may adopt the procedure of replacing K_1 by its maximum term. The maximum term in the summand of K_1 given by (8) occurs at the values of θ_α and θ_β satisfying

$$\log\xi_\alpha+\frac{1}{N_\alpha}\frac{\partial}{\partial\theta_\alpha}\log Q=0,$$
$$\log\xi_\beta+\frac{1}{N_\beta}\frac{\partial}{\partial\theta_\beta}\log Q=0, \tag{12}$$

When K_1 is replaced by its maximum term, Eqs. (9) become identities, Eq. (10) reduces to (6), and Eq. (11) reduces to (5) with the help of (12).

Thus the use of grand partition function is justified.

A more rigorous proof of the above result by means of contour integrals has been given in the paper of the author cited above. This proof will be briefly indicated here with a slight change. Consider the series

$$S(x,y)=\sum x^{N_\alpha\theta_\alpha}y^{N_\beta\theta_\beta}Q(\theta_\alpha,\theta_\beta,T).$$

Thus K_1 given by (8) is $S(\xi_\alpha,\xi_\beta)$ and that given by (7) is $S(\xi,\xi)$. By Cauchy's theorem we have

$$Q(\theta_\alpha,\theta_\beta,T)=\frac{1}{(2\pi i)^2}\oint\oint\frac{S(x,y)}{x^{N_\alpha\theta_\alpha+1}y^{N_\beta\theta_\beta+1}}dxdy.$$

By the general theory of statistical mechanics given in Fowler's book, the integrand has a unique maximum modulus at the saddle point $x=\xi_\alpha$, $y=\xi_\beta$ determined by

$$N_\alpha\theta_\alpha=\xi_\alpha\frac{\partial}{\partial\xi_\alpha}\log S(\xi_\alpha,\xi_\beta),$$
$$N_\beta\theta_\beta=\xi_\beta\frac{\partial}{\partial\xi_\beta}\log S(\xi_\alpha,\xi_\beta).$$

This proves (9). Differentiating Q with respect to θ_α we have

$$\frac{1}{N_\alpha}\frac{\partial Q}{\partial\theta_\alpha}=\frac{1}{(2\pi i)^2}\oint\oint(-\log x)\frac{S(x,y)}{x^{N_\alpha\theta_\alpha+1}y^{N_\beta\theta_\beta+1}}dxdy.$$

There is a similar equation for $\partial Q/\partial\theta_\beta$. Using the method of steepest descent we obtain (12).

To determine the equilibrium value of θ_α for a given value of θ, we put $y=x$ in $S(x,y)$ and obtain

$$\sum{}'Q(\theta_\alpha,\theta_\beta,T)=\frac{1}{2\pi i}\oint\frac{S(x,x)}{x^{N\theta+1}}dx$$

and

$$\sum{}'N_\alpha\theta_\alpha Q(\theta_\alpha,\theta_\beta,T)=\frac{1}{2\pi i}\oint\left[x\frac{\partial}{\partial x}\log S(x,y)\right]_{y=x}\frac{S(x,x)}{x^{N\theta+1}}dx,$$

where the symbol $\sum'$ denotes summation over θ_α with θ_β given as a function of θ_α by (1). Hence the equilibrium value of $N_\alpha\theta_\alpha$ is given by

$$N_\alpha\theta_\alpha=\left[\xi_\alpha\frac{\partial}{\partial\xi_\alpha}\log S(\xi_\alpha,\xi_\beta)\right]_{\xi_\alpha=\xi_\beta=\xi},$$

where ξ is the saddle point determined by

$$N\theta=\xi\frac{\partial}{\partial\xi}\log S(\xi,\xi).$$

This proves (11).

3. APPROXIMATE PARTITION FUNCTION

In Bethe's method one fixes one's attention on a group of quite a small number of sites and constructs the grand partition function for these sites; the effect of omitting other sites of the crystal is taken account of by introducing suitable parameters for the outermost sites of the group. The group of sites are usually divided into the central site, the first shell sites which are nearest neighbors to the central site, the second shell sites which are nearest neighbors to the first shell sites, and so on. In the first approximation the group of sites considered include only the central site and the first shell.

The extension of Bethe's first approximation to include long range interaction has been given in an earlier paper,[6] and the results will be quoted here. The equilibrium values of θ_α and θ_β are given by

$$\frac{\theta_\alpha}{1-\theta_\alpha}=\xi_\alpha \exp(-U_\alpha/kT)\frac{F_\eta}{F},$$
$$\frac{\theta_\beta}{1-\theta_\beta}=\xi_\beta \exp(-U_\beta/kT)\frac{G_\eta}{G}, \tag{13}$$

where F and G are functions of the parameters ϵ_α, ϵ_β, and ζ_α, ζ_β, respectively, and the subscript η attached to a function means that the variables ϵ_α, etc. in that function are changed to $\eta_1\epsilon_\alpha$, etc., where

$$\eta_\nu=\exp(-V_\nu/kT), \qquad \nu=1, 2, \cdots.$$

The parameters ϵ_α, ϵ_β, ζ_α, ζ_β are determined by the following equations:

$$z_{\alpha\alpha}\theta_\alpha=(1-\theta_\alpha)\varphi+\theta_\alpha\varphi_\eta,$$
$$z_{\alpha\beta}\theta_\beta=(1-\theta_\alpha)\psi+\theta_\alpha\psi_\eta, \tag{14}$$
$$z_{\beta\alpha}\theta_\alpha=(1-\theta_\beta)\omega+\theta_\beta\omega_\eta,$$
$$z_{\beta\beta}\theta_\beta=(1-\theta_\beta)\chi+\theta_\beta\chi_\eta, \tag{15}$$

where

$$\varphi=\epsilon_\alpha\frac{\partial}{\partial\epsilon_\alpha}\log F, \quad \psi=\epsilon_\beta\frac{\partial}{\partial\epsilon_\beta}\log F;$$
$$\omega=\zeta_\alpha\frac{\partial}{\partial\zeta_\alpha}\log G, \quad \chi=\zeta_\beta\frac{\partial}{\partial\zeta_\beta}\log G.$$

The quantities U_α and U_β are given by

$$U_\alpha=(U_{\alpha\alpha}-z_{\alpha\alpha}V_1)\theta_\alpha+(U_{\alpha\beta}-z_{\alpha\beta}V_1)\theta_\beta,$$
$$U_\beta=(U_{\beta\alpha}-z_{\beta\alpha}V_1)\theta_\alpha+(U_{\beta\beta}-z_{\beta\beta}V_1)\theta_\beta. \tag{16}$$

The equilibrium value of the energy E_1 is given by

$$E_1=\tfrac{1}{2}N_\alpha\theta_\alpha\{(\varphi_\eta+\psi_\eta)V_1+U_\alpha\}$$
$$+\tfrac{1}{2}N_\beta\theta_\beta\{(\omega_\eta+\chi_\eta)V_1+U_\beta\}. \tag{17}$$

In order to obtain an approximate expression for the partition function Q consistent with Bethe's approximation, it seems that a natural way would be to assume that the equilibrium values of θ_α, θ_β, and E_1 given by (13) and (17) are correct, and to substitute the values of ξ_α and ξ_β given by (13) into (12) and the value of E_1 given by (17) into (6) to obtain Q by integration. Unfortunately this procedure does not work, because the system of differential equations for Q so obtained does not, in general, satisfy the condition of integrability.

In my previous paper this difficulty was avoided by regarding U_α and U_β as unknown functions to be determined by the condition of integrability. A system of partial differential equations for these two quantities was thus obtained, whose solutions were made unique by the condition that U_α and U_β reduce to the expressions given by (16) in the special case $\theta_\alpha=\theta_\beta$ and in the special case $\theta=1$. These conditions are evidently obtained from physical requirement.

The physical basis for the above procedure is that the assumption of a uniform distribution of A atoms with a fraction θ_α on α-sites and a fraction θ_β on β-sites (outside the small group of sites) used in deriving (16) is not true and should be replaced by the condition of integrability. But the assumption is certainly true in the special cases $\theta_\alpha=\theta_\beta$ and $\theta=1$, and so (16) serves as a boundary condition.

It has been mentioned in the Introduction that the above theory suffers from two defects. The most serious defect is that in the special case of neighbor interaction the two quantities U_α and U_β should be zero, but then the condition of integrability is not satisfied in general. If we again introduce U_α and U_β into the formula to restore the condition of integrability, it is hard to interpret their physical significance. A second defect is that the numerical work involved in calculating U_α and U_β is too laborious.

We shall now modify the theory by preserving formulae (16) and (17) for U_α, U_β, and E_1, but regarding ξ_α and ξ_β given by (13) as incorrect. *We shall obtain Q by integrating (6), and take (12) as the equations determining ξ_α and ξ_β.* To integrate (6) we remember that as $T\to\infty$, Q approaches the value

$$g=g(\theta_\alpha, \theta_\beta)=\binom{N_\alpha}{N_\alpha\theta_\alpha}\binom{N_\beta}{N_\beta\theta_\beta}. \tag{18}$$

Hence

$$\log Q=\log g-\int_0^{1/T}\frac{E_1}{k}d\left(\frac{1}{T}\right), \tag{19}$$

in which the value of E_1 given by (17) should be used.

[6] J. S. Wang, Proc. Roy. Soc. **A168**, 56 (1938).

When Q has been obtained in this way, the equilibrium values of θ_α and θ_β will be given by (5), instead of by (13) and (11).

We shall now transform the integral in (19) to a simpler form. Let us denote the local grand partition function for central α-site by K_α, and write λ_α in place of $\xi_\alpha \exp(-U_\alpha/kT)$, remembering that λ_α has no real connection with ξ_α in (12) owing to the breakdown of the integrability condition. Similarly, we have K_β and λ_β for central β-site. Thus

$$K_\alpha = F + \lambda_\alpha F_\eta, \quad K_\beta = G + \lambda_\beta G_\eta. \tag{20}$$

The quantities λ_α, ϵ_α, etc., are determined as functions of θ_α, θ_β, T by (13)–(15). As $T \to \infty$, these quantities reduce to

$$\lambda_\alpha = \epsilon_\alpha = \zeta_\alpha = \frac{\theta_\alpha}{1-\theta_\alpha}, \quad \lambda_\beta = \epsilon_\beta = \zeta_\beta = \frac{\theta_\beta}{1-\theta_\beta}. \tag{21}$$

Then the result of integrating (19) is

$$\begin{aligned}\log Q = \log g &- \tfrac{1}{2}(N_\alpha\theta_\alpha U_\alpha + N_\beta\theta_\beta U_\beta)/kT + \tfrac{1}{2}N_\alpha\{\log K_\alpha - \theta_\alpha \log \lambda_\alpha - z_{\alpha\alpha}\theta_\alpha \log \epsilon_\alpha - z_{\alpha\beta}\theta_\beta \log \epsilon_\beta\} \\ &+ \tfrac{1}{2}N_\beta\{\log K_\beta - \theta_\beta \log \lambda_\beta - z_{\beta\alpha}\theta_\alpha \log \zeta_\alpha - z_{\beta\beta}\theta_\beta \log \zeta_\beta\} \\ &+ \tfrac{1}{2}(z+1)[N_\alpha\theta_\alpha \log \theta_\alpha + N_\alpha(1-\theta_\alpha)\log(1-\theta_\alpha) + N_\beta\theta_\beta \log \theta_\beta + N_\beta(1-\theta_\beta)\log(1-\theta_\beta)] \\ &+ \tfrac{1}{2}\int_0^{1/T} \{N_\alpha(1-\theta_\alpha)F^* + N_\alpha\theta_\alpha F_\eta^* + N_\beta(1-\theta_\beta)G^* + N_\beta\theta_\beta G_\eta^*\}\frac{V_1}{k} d\left(\frac{1}{T}\right),\end{aligned} \tag{22}$$

where F^* is defined by

$$F^* = \frac{kT^2}{V_1}\frac{\partial}{\partial T}\log F(\epsilon_\alpha, \epsilon_\beta). \tag{23}$$

Since it follows from Stirling's formula that

$$\log g = -N_\alpha[\theta_\alpha \log \theta_\alpha + (1-\theta_\alpha)\log(1-\theta_\alpha)] - N_\beta[\theta_\beta \log \theta_\beta + (1-\theta_\beta)\log(1-\theta_\beta)], \tag{24}$$

we have

$$\begin{aligned}\log Q = \left(1-\frac{z}{2}\right)\log g &- \tfrac{1}{2}(N_\alpha\theta_\alpha U_\alpha + N_\beta\theta_\beta U_\beta)/kT + \tfrac{1}{2}N_\alpha[\theta_\alpha \log F_\eta + (1-\theta_\alpha)\log F \\ &- z_{\alpha\alpha}\theta_\alpha \log \epsilon_\alpha - z_{\alpha\beta}\theta_\beta \log \epsilon_\beta] + \tfrac{1}{2}N_\beta[\theta_\beta \log G_\eta + (1-\theta_\beta)\log G - z_{\beta\alpha}\theta_\alpha \log \zeta_\alpha - z_{\beta\beta}\theta_\beta \log \zeta_\beta] \\ &+ \tfrac{1}{2}\int_{\eta_1}^1 \{N_\alpha(1-\theta_\alpha)F^* + N_\alpha\theta_\alpha F_\eta^* + N_\beta(1-\theta_\beta)G^* + N_\beta\theta_\beta G_\eta^*\}\frac{d\eta_1}{\eta_1}.\end{aligned} \tag{25}$$

The last integral should be evaluated by numerical integration.

4. HIGHER APPROXIMATIONS

The method of obtaining the partition function in the first approximation given in the preceding section is evidently applicable to higher approximations. The group of sites considered in the general case is divided into the central site, the first shell sites which are at a distance a_1 from the central site, the sites which are at a distance a_2 from the central site, and so on. Let the numbers of α- and β-sites among those which are at a distance a_r from the central site be $z_{r\alpha\alpha}$, $z_{r\alpha\beta}$ if the central site is an α, and $z_{r\beta\alpha}$, $z_{r\beta\beta}$ if the central site is a β. These numbers satisfy the relations

$$z_{r\alpha\alpha} + z_{r\alpha\beta} = z_{r\beta\alpha} + z_{r\beta\beta} = z_r. \tag{26}$$

Let us consider the group with a central α-site. We shall introduce a parameter λ_α for an A atom on the central site, and parameters $\epsilon_{r\alpha}$ and $\epsilon_{r\beta}$, respectively, for an A atom on the α- and β-sites which are at a distance a_r from the central site. In any distribution of A atoms in the group of sites considered, let there be n_0 A atoms in the central site ($n_0 = 1$ or 0 according as the central site is occupied or not), $n_r = n_r' + n_r''$ A atoms in the sites at a distance a_r from the central site, n_r' being the number in the α-sites and n_r'' the number in the β-sites of them. Let m_r be the number of pairs of A atoms in the

group at a distance a_ν apart, and let γ_α be the number of arrangements for a given set of values n_ν', n_ν'', and m_ν. Then the local grand partition function is

$$K_\alpha = F + \lambda_\alpha F_\eta,$$

with

$$F = F(\epsilon_{1\alpha}, \epsilon_{1\beta}, \epsilon_{2\alpha}, \epsilon_{2\beta}, \cdots) = \sum \gamma_\alpha \prod_{\nu=1}^{\kappa} (\epsilon_{\nu\alpha}^{n_\nu'} \epsilon_{\nu\beta}^{n_\nu''} \eta_\nu^{m_\nu'}) \prod_{\nu>\kappa} \eta_\nu^{m_\nu},$$

$$F_\eta = F(\eta_1\epsilon_{1\alpha}, \eta_1\epsilon_{1\beta}, \eta_2\epsilon_{2\alpha}, \eta_2\epsilon_{2\beta}, \cdots), \tag{27}$$

in which the suffix κ denotes the outermost sites of the group and the relations

$$m_\nu = n_0 n_\nu + m_\nu', \quad (\nu \leqq \kappa)$$

have been used.

The parameters will be determined by the conditions

$$\theta_\alpha = \lambda_\alpha \frac{\partial}{\partial \lambda_\alpha} \log K_\alpha, \tag{28}$$

$$z_{\nu\alpha\alpha}\theta_\alpha = \epsilon_{\nu\alpha}\frac{\partial}{\partial \epsilon_{\nu\alpha}} \log K_\alpha, \quad z_{\nu\alpha\beta}\theta_\beta = \epsilon_{\nu\beta}\frac{\partial}{\partial \epsilon_{\nu\beta}} \log K_\alpha. \tag{28$_\nu$}$$

These equations will also be expressed in a new notation, analogous to (14),

$$z_{\nu\alpha\alpha}\theta_\alpha = (1-\theta_\alpha)\varphi_\nu + \theta_\alpha\varphi_{\nu\eta}, \quad z_{\nu\alpha\beta}\theta_\beta = (1-\theta_\alpha)\psi_\nu + \theta_\alpha\psi_{\nu\eta}, \tag{29}$$

where

$$\varphi_\nu = \epsilon_{\nu\alpha}\frac{\partial}{\partial \epsilon_{\nu\alpha}} \log F, \quad \psi_\nu = \epsilon_{\nu\beta}\frac{\partial}{\partial \epsilon_{\nu\beta}} \log F.$$

There are a similar set of equations for the group with central β-site.

The present formulation of the higher approximations is different from Bethe's original formulation and from the modified one given by the author[7] for the case of adsorption with long range interaction. Bethe considered only the special case of equal A and B atoms and treated them symmetrically, so that in the first approximation he put $\lambda_\alpha = 1$ and in the second approximation he put $\lambda_\alpha = 1$, $\epsilon_{1\beta} = 1$ (in Bethe's case there are no α-sites in the first shell surrounding a central α-site). When the numbers of A and B atoms are not equal, it is more convenient to concentrate attention solely on A atoms and to introduce a λ_α different from unity, as first done by Peierls.[8] Such a procedure also works in the case of equal A and B atoms.[9] Peierls put $\lambda_\alpha = \xi_\alpha = \xi$ as the condition of equilibrium. This was also followed by the author[6,9] in the case of long range interaction by putting $\lambda_\alpha = \xi \exp(-U_\alpha/kT)$, the second factor being introduced to take account of the influence of exterior sites. Since this does not lead to consistent results when the partition function is sought, we have replaced it by Eq. (5) in the preceding section. In accordance with a procedure adopted in the theory of adsorption with long range interaction,[7] a natural generalization to higher approximations would be to put

$$\epsilon_{\nu\alpha} = \xi_\alpha \exp(-U_{\nu\alpha}'/kT), \quad \epsilon_{\nu\beta} = \xi_\beta \exp(-U_{\nu\alpha}''/kT), \tag{30}$$

where $U_{\nu\alpha}'$ is the energy owing to exterior sites on an α-site at a distance a_ν from a central α-site, and $U_{\nu\alpha}''$ is the energy owing to exterior sites on a β-site at a distance a_ν from a central α-site. Such a generalization evidently fails for the same reason as Peierls' assumption on λ_α does. Therefore we replace it by the condition (28$_\nu$).

---

[7] J. S. Wang, Proc. Camb. Phil. Soc. **34**, 238 (1938).
[8] R. Peierls, Proc. Roy. Soc. **A154**, 207 (1936).
[9] See J. S. Wang, Proc. Roy. Soc. **A168**, 68 (1938).

The approximate partition function will still be given by (19), in which the energy now has the form

$$E_1=\tfrac{1}{2}N_\alpha\theta_\alpha\left\{\sum_{\nu=1}^{\kappa}(\varphi_{\nu\eta}+\psi_{\nu\eta})V_\nu+U_\alpha\right\}+\tfrac{1}{2}N_\beta\theta_\beta\left\{\sum_{\nu=1}^{\kappa}(\omega_{\nu\eta}+\chi_{\nu\eta})V_\nu+U_\beta\right\}, \tag{31}$$

where U_α and U_β are given by

$$\begin{aligned}U_\alpha&=\left(U_{\alpha\alpha}-\sum_{\nu=1}^{\kappa}z_{\nu\alpha\alpha}V_\nu\right)\theta_\alpha+\left(U_{\alpha\beta}-\sum_{\nu=1}^{\kappa}z_{\nu\alpha\beta}V_\nu\right)\theta_\beta,\\ U_\beta&=\left(U_{\beta\alpha}-\sum_{\nu=1}^{\kappa}z_{\nu\beta\alpha}V_\nu\right)\theta_\alpha+\left(U_{\beta\beta}-\sum_{\nu=1}^{\kappa}z_{\nu\beta\beta}V_\nu\right)\theta_\beta.\end{aligned} \tag{32}$$

If we introduce the notation

$$F^*=\frac{kT^2}{V_1}\frac{\partial}{\partial T}\log F(\epsilon_{1\alpha},\epsilon_{1\beta},\epsilon_{2\alpha},\epsilon_{2\beta},\cdots)=\frac{1}{FV_1}\sum\gamma_\alpha\left(\sum_{\nu=1}^{\kappa}m_\nu'V_\nu+\sum_{\nu>\kappa}m_\nu V_\nu\right)\prod_{\nu=1}^{\kappa}(\epsilon_{\nu\alpha}^{n_\nu'}\epsilon_{\nu\beta}^{n_\nu''}\eta_\nu^{m_\nu'})\prod_{\nu>\kappa}\eta_\nu^{m_\nu} \tag{33}$$

similar to (23), then

$$\begin{aligned}\log Q=\log g&-\tfrac{1}{2}\left(1+\sum_{\nu=1}^{\kappa}z_\nu\right)\log g-\tfrac{1}{2}(N_\alpha\theta_\alpha U_\alpha+N_\beta\theta_\beta U_\beta)/kT\\ &+\tfrac{1}{2}N_\alpha\left\{\log K_\alpha-\theta_\alpha\log\lambda_\alpha-\sum_{\nu=1}^{\kappa}z_{\nu\alpha\alpha}\theta_\alpha\log\epsilon_{\nu\alpha}-\sum z_{\nu\alpha\beta}\theta_\beta\log\epsilon_{\nu\beta}\right\}\\ &+\tfrac{1}{2}N_\beta\{\log K_\beta-\theta_\beta\log\lambda_\beta-\sum z_{\nu\beta\alpha}\theta_\alpha\log\zeta_{\nu\alpha}-\sum z_{\nu\beta\beta}\theta_\beta\log\zeta_{\nu\beta}\}\\ &+\tfrac{1}{2}\int_0^{1/T}\{N_\alpha(1-\theta_\alpha)F^*+N_\alpha\theta_\alpha F_\eta^*+N_\beta(1-\theta_\beta)G^*+N_\beta\theta_\beta G_\eta^*\}\frac{V_1}{k}\alpha\left(\frac{1}{T}\right).\end{aligned} \tag{34}$$

The actual calculation may sometimes be simplified by assuming reasonable values for some of the $\epsilon_{\nu\alpha}$, etc., and dropping the corresponding equations in (28_ν). Thus in the case of the second approximation, for example, one may assume $\epsilon_{1\alpha}$ and $\epsilon_{1\beta}$ to be given by (30), with ξ_α, ξ_β determined by the first approximation, and drop the corresponding equation (28_1). In a certain sense, it is better to drop the equation (28_2) and retain (28_1) to determine $\epsilon_{2\alpha}$ and $\epsilon_{2\beta}$. This alternative procedure has the advantage in making the approximation better for the central site and the first shell, and actually corresponds to Bethe's original procedure. Bethe also assumed one and the same parameter for A atoms in the corner and medium sites of the second shell. This procedure has the disadvantage of losing the symmetry, and it is perhaps still better to retain all the equations (28_ν), in which the corner and medium sites are distinguished by the difference in ν, i.e., in their distances from the central site.

A different way of calculating the energy has been adopted by Mr. Yang in his generalization of the quasi-chemical method. His method can also be slightly modified to apply to the present problem. In obtaining the expression (31) for the energy we have assumed that the equilibrium values of M_ν^{AA} for $\nu\leqq\kappa$ are given by

$$\tfrac{1}{2}N_\alpha\langle(n_0n_\nu)_\alpha\rangle_M+\tfrac{1}{2}N_\beta\langle(n_0n_\nu)_\beta\rangle_M.$$

We shall now modify this assumption for $\nu=1$ but shall retain the assumption for $\nu>1$. In the case of $\nu=1$, let us assume that m_1 is the same fraction of μ as $\langle M_1^{AA}\rangle_M$ is of $\frac{1}{2}Z_1=\frac{1}{2}Nz$, where μ and $\frac{1}{2}Nz$ are the total numbers of pairs of neighboring sites for the group and for the whole crystal, respectively.[10]

[10] The value of μ is the same for central α- and for central β-site on account of equal z for them.

Under this assumption, we have

$$kT^2\frac{\partial}{\partial T}\log K_\alpha=\frac{1}{K_\alpha}\sum\gamma_\alpha(\sum m_\nu V_\nu)\lambda_\alpha{}^{n_0}\prod_{\nu=1}^{\kappa}\epsilon_{\nu\alpha}{}^{n_\nu'}\epsilon_{\nu\beta}{}^{n_\nu''}\prod_\nu\eta_\nu{}^{m_\nu}$$

$$=\sum\bar{m}_\nu V_\nu=\frac{2\mu}{Nz}\langle M_1{}^{AA}\rangle_{kl}V_1+\sum_{\nu>1}\bar{m}_\nu V_\nu.$$

Summing over all central α-groups and all central β-groups we obtain

$$N_\alpha kT^2\frac{\partial}{\partial T}\log K_\alpha+N_\beta kT^2\frac{\partial}{\partial T}\log K_\beta=\frac{2\mu}{z}\langle M_1{}^{AA}\rangle_{kl}V_1+N_\alpha(\sum_{\nu>1}\bar{m}_\nu V_\nu)_\alpha+N_\beta(\sum_{\nu>1}\bar{m}_\nu V_\nu)_\beta.$$

Hence the modified energy expression is

$$E_1=\frac{z}{2\mu}kT^2\frac{\partial}{\partial T}(N_\alpha\log K_\alpha+N_\beta\log K_\beta)-\frac{z}{2\mu}\{N_\alpha(\sum_{\nu>1}\bar{m}_\nu V_\nu)_\alpha+N_\beta(\sum_{\nu>1}\bar{m}_\nu V_\nu)_\beta\}$$

$$+\tfrac{1}{2}N_\alpha\theta_\alpha\left\{\sum_{\nu=2}^{\kappa}(\varphi_{\nu\eta}+\psi_{\nu\eta})V_\nu+U_\alpha\right\}+\tfrac{1}{2}N_\beta\theta_\beta\left\{\sum_{\nu=2}^{\kappa}(\omega_{\nu\eta}+\chi_{\nu\eta})V_\nu+U_\beta\right\}. \quad (35)$$

When this expression is substituted into (19), the partition function becomes, after integration,

$$\log Q=\log g-\frac{z}{2\mu}(1+\sum z_\nu)\log g-\tfrac{1}{2}(N_\alpha\theta_\alpha U_\alpha+N_\beta\theta_\beta U_\beta)/kT$$

$$+\frac{z}{2\mu}N_\alpha\{\log K_\alpha-\theta_\alpha\log\lambda_\alpha-\sum z_{\nu\alpha\alpha}\theta_\alpha\log\epsilon_{\nu\alpha}-\sum z_{\nu\alpha\beta}\theta_\beta\log\epsilon_{\nu\beta}\}$$

$$+\frac{z}{2\mu}N_\beta\{\log K_\beta-\theta_\beta\log\lambda_\beta-\sum z_{\nu\beta\alpha}\theta_\alpha\log\zeta_{\nu\alpha}-\sum z_{\nu\beta\beta}\theta_\beta\log\zeta_{\nu\beta}\}$$

$$+\frac{z}{2\mu k}\sum_{\nu>1}V_\nu\int_0^{1/T}\{N_\alpha(\bar{m}_\nu)_\alpha+N_\beta(\bar{m}_\nu)_\beta\}\,d\left(\frac{1}{T}\right)$$

$$-\frac{1}{2}\sum_{\nu=2}^{\kappa}\frac{V_\nu}{k}\int_0^{1/T}\{N_\alpha\theta_\alpha(\varphi_{\nu\eta}+\psi_{\nu\eta})+N_\beta\theta_\beta(\omega_{\nu\eta}+\chi_{\nu\eta})\}d\left(\frac{1}{T}\right). \quad (36)$$

The advantage of this alternative formulation in the case of neighbor interaction is evident. For then the integral in (36) is identically zero and the partition function is obtained as a closed expression. In the particular case of first approximation for the simple cubic and body-centered cubic lattices these two formulations are the same.

It may be argued that since Bethe's method would give a better equilibrium value for the central site than for the outer sites, the expression (31) for the energy is perhaps better than (35).

5. THE PROBLEM OF ADSORPTION

In the problem of adsorption with a single kind of adsorbed atoms, the theory of superlattices can be applied provided the A atoms are interpreted as the adsorbed atoms. It is now more convenient to choose θ, s, T as independent variables, the relations between θ, s and θ_α, θ_β are given by (1) and (2). The equilibrium value of s as a function of θ and T is still given by (5). But θ depends on the pressure p of the gas, and so instead of regarding (12) as equations determining ξ_α and ξ_β in terms of θ and s, we regard them as determining θ in terms of ξ and T, with ξ connected with the pressure by the relation:[11]

$$\xi=\frac{h^3e^{\chi/kT}}{(2\pi m)^{3/2}(kT)^{5/2}}\frac{v_s(T)}{b_g(T)}p. \quad (37)$$

[11] See J. S. Wang, Proc. Roy. Soc. **A161**, 127 (1937).

In terms of θ and s, each one of Eq. (12) becomes

$$\log \xi = -\frac{1}{N}\frac{\partial}{\partial \theta}\log Q. \tag{38}$$

This equation gives the relation between θ, p, and T, and is commonly known as the adsorption isotherm.

The heat of adsorption q is defined as the decrease of energy when one atom is transferred from the gas phase to the adsorbed phase and is given by[11]

$$\begin{aligned} q &= q_0 - \frac{1}{N}\frac{\partial E_1}{\partial \theta}, \\ q_0 &= \chi + \tfrac{3}{2}kT + kT^2\frac{d}{dT}\log\frac{b_g}{v_a}, \end{aligned} \tag{39}$$

where E_1 is considered as a function of θ and T and the equilibrium value of s determined by (5) should be substituted in it.

The special case of equal numbers of α- and β-sites is interesting for its simplicity. It is easy to see that in this case E_1 is symmetrical in α and β, and so it is an even function of s. To study the nature of the dependence of E_1 on θ, we shall consider the first approximation for simplicity. It will be evident that the higher approximations can be treated in the same way. In the case of first approximation, there is a single ϵ and a single ζ. The functions ψ and ω satisfy the relation[9]

$$f(x) = z - f(\eta_1^{2\mu+1}x^{-1}), \quad (f=\psi \text{ or } \omega). \tag{40}$$

It then follows from (14) and (15) that the transformation

$$\epsilon \to \eta_1^{2\mu}\zeta^{-1}, \quad \zeta \to \eta_1^{2\mu}\epsilon^{-1} \tag{41}$$

changes θ_α to $1-\theta_\beta$ and θ_β to $1-\theta_\alpha$, so that s is unaltered while θ is changed to $1-\theta$. Then (17) shows that

$$E_1 = \tfrac{1}{2}NU(\theta - \tfrac{1}{2}), \quad (U = U_{\alpha\alpha} + U_{\alpha\beta}),$$

is an even function of $\theta-\frac{1}{2}$. Consequently

$$\log \xi - \frac{U}{2kT} \quad \text{and} \quad q - q_0 + \tfrac{1}{2}U$$

are both odd functions of $\theta-\frac{1}{2}$ and even functions of s.

The same conclusion holds in the case of vanishing order ($s=0$) in a general lattice (i.e., r need not be $\frac{1}{2}$). For then $\theta_\alpha = \theta_\beta = \theta$, $\epsilon_\alpha = \epsilon_\beta = \zeta_\alpha = \zeta_\beta = \epsilon$, $F(\epsilon, \epsilon) = G(\epsilon, \epsilon) = f(\epsilon)$, and E_1 in the first approximation becomes

$$E_1 = \tfrac{1}{2}N\theta\{\psi_1 V_1 + (U - zV_1)\theta\}. \tag{42}$$

It is easily seen that $E_1 - \frac{1}{2}NU(\theta-\frac{1}{2})$ is an even function of $\theta-\frac{1}{2}$, and so the same conclusion on ξ and q follows.

PHYSICAL REVIEW VOLUME 67, NUMBERS 3 AND 4 FEBRUARY 1 AND 15, 1945

The Multiple Scattering of Waves*

I. General Theory of Isotropic Scattering by Randomly Distributed Scatterers

LESLIE L. FOLDY**
Columbia University, New York, New York
(Received November 29, 1944)

While the problem of the multiple scattering of particles by a random distribution of scatterers has been treated classically through the use of the Boltzmann integro-differential equation, the corresponding problem of the multiple scattering of waves seems to have received scant attention. All previous treatments have considered the problem in the "geometrical optics" limit, where the rays are regarded as trajectories of particles and the treatment for particles is then applied, so that the interference phenomena in wave scattering are neglected. In this paper the problem of the multiple scattering of scalar waves by a random distribution of isotropic scatterers is considered in detail on the basis of a consistent wave treatment. The introduction of the concept of "randomness" requires averages to be taken over a statistical ensemble of scatterer configurations. Equations are derived for the average value of the wave function, the average value of the square of its absolute value, and the average flux carried by the wave. The second of these quantities satisfies an integral equation which has some similarities to the corresponding equation for particle scattering. The physical interpretation of the results is discussed in some detail and possible generalizations of the theory are outlined.

INTRODUCTION

THE problem of the multiple scattering of particles by a random distribution of scatterers has been treated in some detail in literature.[1] These treatments are all based essentially on the Boltzmann integro-differential equation of conservation in phase space and are thus classical rather than quantum-mechanical in foundation. On the other hand, a consistent treatment, based on the wave equation itself, of the multiple scattering of waves by a randomly distributed collection of scatterers, has not, to the writer's knowledge, yet appeared in literature. Those treatments which have appeared, for example, in discussions of the problem of the propagation of light through stellar atmospheres and through turbid media,[2] are all based on the approximation of considering light rays as the trajectories of particles (photons), which are then treated on the basis of the Boltzmann equation. This, in a sense, corresponds to treating light scattering in the "geometrical optics limit" where interference phenomena characteristic of waves is neglected. Such a treatment might be expected, *a priori*, to be valid if the wave-length of the waves is small compared to the mean free path of the associated particles in the scattering medium, but the only satisfactory justification for such a treatment lies in showing that the wave treatment leads to the same results.

The investigation of the multiple scattering of waves is not only important for the above reason but also for several others. The multiple scattering of sound waves by the water droplets of a fog or by other collections of small obstacles often occurs in cases where the interference effects may not be legitimately neglected, and thus this problem cannot be satisfactorily discussed except by a wave treatment. Also in the scattering of electrons and other fundamental particles by aggregates of nuclei, atoms, or molecules in gases, liquids, and amorphous solids, the use of the classical particle mechanics may be questionable, and the quantum-mechanical treatment of this problem necessitates the inves-

* Publication assisted by Ernest Kempton Adams Fund for Physical Research of Columbia University.
** Correspondence should be addressed to 64th Floor, 350 Fifth Avenue, New York 1, New York.
[1] E. Fermi, Ricerca Scienta **VII–II**, 13 (1936); Yost and Dickinson, Phys. Rev. **50**, 128 (1936); G. C. Wick, Atti del Acad. Reale dei Lincei **23**, 775 (1936); L. S. Ornstein, Kon. Akad. van Wet. te Amst. Proc. **XXXIX**, No. 9 (1936); Halpern, Lueneburg, and Clark, Phys. Rev. **53**, 173 (1938); Schuchard and Uehling, Phys. Rev. **58**, 611 (1940).
[2] E. Hopf, "Problems of radiative equilibrium," Cambridge Tract No. 31 (1934); E. A. Milne, "Thermodynamics of stars," *Handbuch der Astrophys.*, Vol. III/I (Julius Springer, Berlin, 1930); G. Jaffe, Ann. d. Physik [5] **6**, 195 (1930); S. Q. Duntley, J. Opt. Soc. Am. **33**, 252 (1934); E. O. Hulbert, J. Opt. Soc. Am. **33**, 42 (1943).

tigation of the multiple scattering of waves. Finally, the connection between the so-called coherent scattering (leading to the macroscopic phenomena of refraction and specular reflection) and the incoherent scattering can only be appreciated through a consistent wave treatment of scattering from beginning to end.

In this paper we shall develop a wave theory of multiple scattering for the case of scalar waves which are isotropically scattered by a random distribution of scatterers, the "randomness" being defined explicitly in the next section. We deal with abstract waves of unspecified type so that the treatment is applicable to any type of wave satisfying the scalar wave equation. The concept of randomness being essentially statistical, it is necessary to concentrate one's interest in the average values of physical quantities, the averages being taken over a statistical ensemble of collections of scatterers. The quantities of greatest interest are the wave function itself, the square of its absolute magnitude, or its mean square value, and the flux of some quantity (energy, probability, etc.) carried by the wave.

The determination of the average value of the wave function is practically identical with the already treated theory of the index of refraction of material media (other than crystalline media) for light,[3] but is included in the present treatment since it forms a necessary introduction to the second problem of determining the average value of the square of the absolute value of the wave function. It is shown that the average value of the wave function satisfies the wave equation in a continuous medium in which no scatterers are present and in which the velocity of propagation is different from the velocity in the original medium in the absence of the scatterers and is, in general, a function of position and complex rather than real in value. In other words, the incident wave and the scattered waves combine on the average to form a wave which travels uniformly without scattering at a different velocity from the incident wave and with attenuation. This wave shows refraction and reflection phenomena and thus describes the coherent aspects of scattering, such as the specular reflection taking place at a boundary between a region where scatterers are present (in sufficiently great density) and a region where no scatterers are present. An example would be the reflection of light at a glass-vacuum interface, the atoms of the glass being the scatterers.

The principal contribution of this paper, however, lies in the treatment of the average of the square of the wave function, which is in general different from the square of the average value of the wave function. The difference represents what is usually referred to as the incoherent scattering, the term incoherent referring not to a lack of definite phase relationship between the incident wave and the wave scattered by a particular scatterer, but to the statistical superposition of these scattered waves when propagated to the point of observation because of the "random" distribution of the scatterers. The average of the square of the wave function will be shown to satisfy an integral equation which has some similarities to the form of the Boltzmann integro-differential equation governing the multiple scattering of particles. This integral equation is just as fundamental to the problem of the multiple scattering of waves as the Boltzmann integro-differential equation is to the problem of the multiple scattering of particles. The physical interpretation of the integral equation may be fairly readily recognized: The equation states that the average of the square of the wave function at any point is equal to the square of the average wave function at the point plus contributions representing scattering from all other portions of the medium; these contributions are proportional to the average of the square of the wave function at each point and to a scattering cross section per unit volume at each of these points. The value of the scattering cross section per unit volume at each point and the mode of propagation from each point to the point of interest are exactly defined in terms of the scattering properties of the individual scatterers and the properties of the statistical ensemble of scatterers studied. In fact the integral equation could be derived from its *a posteriori* physical interpretation, but in that case the value of the scattering cross section per unit volume and the mode of propagation of the scattered waves would have to be intro-

[3] M. Born, *Optik* (Julius Springer, Berlin, 1933), pp. 313 ff.

duced *ad hoc*, so that such a derivation could hardly be considered rigorous or fundamental.

STATISTICAL CONSIDERATIONS

Since there is no inherent limitation in the theory restricting it to the case where all scatterers have identical scattering properties, we shall assume that the scattering properties of the scatterers are determined by a *scattering parameter* denoted by s. In some applications s may be a continuously varying physical parameter such as the radius of water droplets in the problem of the scattering of sound by a fog, or it may be simply an index identifying particular types of scatterers such as various types of nuclei present in the problem of the scattering of neutrons by an amorphous solid or by a gas. In what follows we shall assume it to be a continuous variable, but the transformation of the equations in case it takes on only discrete values involves only an obvious replacement of integrals by summations.

If we have a collection of N scatterers and are given for each its position and the value of s which characterizes its scattering properties we shall say that we have a particular configuration of the scatterers. Now, in the problem of multiple scattering of waves by a collection of amorphously distributed scatterers, one rarely has sufficient information to establish the microscopic configuration of the collection and one is therefore not so much interested in the values of physical quantities for a particular configuration of the scatterers as in the average value of these quantities, the average being taken over the ensemble of possible configurations of the scatterers consistent with what information one has as to the macroscopic state of the collection. Thus if we specify the positions of the scatterers by their position vectors $\mathbf{r}_1, \mathbf{r}_2, \cdots \mathbf{r}_N$, and their scattering properties by the values of the scattering parameters $s_1, s_2, \cdots s_N$, a particular configuration of the scatterers will be given when all the $\mathbf{r}_j$'s and s_j's are specified. We may then indicate the ensemble of configurations in which we are interested by a probability distribution function $P(\mathbf{r}_1, \mathbf{r}_2, \cdots \mathbf{r}_N, s_1, s_2, \cdots s_N)$ so that

$$P(\mathbf{r}_1, \mathbf{r}_2, \cdots \mathbf{r}_N, s_1, s_2, \cdots s_N)d\mathbf{r}_1 d\mathbf{r}_2 \cdots \times d\mathbf{r}_N ds_1 ds_2 \cdots ds_N, \quad (1)$$

represents the probability of finding the scatterers in a configuration in which the first scatterer lies in the element of volume $d\mathbf{r}_1$ about the point $\mathbf{r}_1$ and has a scattering parameter lying between s_1 and s_1+ds_1, the second scatterer lies in the element of volume $d\mathbf{r}_2$ about the point $\mathbf{r}_2$ and has a scattering parameter lying between s_2 and s_2+ds_2, etc. To find the average of a physical quantity over the ensemble of configurations we then multiply its value for the configuration $(\mathbf{r}_1, \mathbf{r}_2, \cdots \mathbf{r}_N, s_1, s_2, \cdots s_N)$ by (1) and integrate over all values of the $\mathbf{r}_j$'s and s_j's accessible to the scatterers.

In this paper we shall be interested only in probability distribution functions in which the probability that a particular scatterer is located in some volume element and has a value of s located within some range ds is independent of the locations and scattering parameters of the other scatterers and is the same for all scatterers. In this case we may write for the probability distribution function

$$P=(1/N)^N n(\mathbf{r}_1, s_1)n(\mathbf{r}_2, s_2)\cdots n(\mathbf{r}_N, s_N). \quad (2)$$

The average number of scatterers per unit volume in the neighborhood of the point $\mathbf{r}$ having scattering parameters lying between s and $s+ds$ is then

$$n(\mathbf{r}, s)ds,$$

and the total number of scatterers per unit volume in the neighborhood of the point $\mathbf{r}$ is

$$n(\mathbf{r})=\int n(\mathbf{r}, s)ds,$$

the integration being taken over all values of s. Finally the normalization of the probability requires

$$\int_V\int n(\mathbf{r}, s)ds d\mathbf{r}=N$$

where V represents the total volume accessible to the scatterers.

We shall call the average of a physical quantity over the ensemble of configurations a *configurational average* and shall denote it by enclosing the quantity which is averaged in angular brackets. Thus if f is a function of the $\mathbf{r}_j$'s and s_j's, its

configurational average is

$$\langle f\rangle=\int_V\cdots\int_V\int_V\cdots\int f(\mathbf{r}_j, s_j)\times P(\mathbf{r}_j, s_j)ds_1ds_2\cdots ds_N d\mathbf{r}_1d\mathbf{r}_2\cdots d\mathbf{r}_N. \quad (3)$$

If we wish to omit the integration over the positions and scattering powers of one or more of the scatterers we shall indicate this by subscribing the indices of these scatterers to the angular brackets; thus $\langle f\rangle_j$ indicates that the integration over $\mathbf{r}_j$ and s_j is to be omitted in (3).

STATEMENT OF THE PROBLEM

In this section we shall make a precise statement of the problem which we shall treat in this paper. Interest here will be restricted to the case of steady-state scattering of waves of a single frequency ω so that the value of the scalar wave function at the point $\mathbf{r}$ at time t can be represented as $\psi(\mathbf{r})e^{i\omega t}$. In the absence of scatterers from the medium, $\psi(\mathbf{r})$ will then satisfy the wave equation

$$\nabla^2\psi+k_0^2\psi=0, \quad (4)$$

where $k_0=\omega/c_0$ and c_0 is the wave velocity in the scatterer-free medium. Furthermore we shall consider here only isotropic point scatterers, that is, those which scatter spherically symmetrical waves (s-wave scattering only). In this case the wave function in the presence of the scatterers will have simple poles at the positions of the scatterers and in the neighborhood of the jth scatterer will behave like

$$\frac{A_j\exp[-ik_0|\mathbf{r}-\mathbf{r}_j|]}{|\mathbf{r}-\mathbf{r}_j|}.$$

We may then define the *external field* acting on the jth scatterer as

$$\psi^j(\mathbf{r})=\psi(\mathbf{r})-\frac{A_j\exp[-ik_0|\mathbf{r}-\mathbf{r}_j|]}{|\mathbf{r}-\mathbf{r}_j|}, \quad (5)$$

this field having no singularity at $\mathbf{r}=\mathbf{r}_j$. We then characterize the scattering properties of the scatterers by the relationship

$$A_j=g(s_j, \omega)\psi^j(\mathbf{r}_j), \quad (6)$$

making the strength of the scattered wave from a scatterer proportional to the external field acting on it. In the above $g(s_j, \omega)$ will be referred to as the *scattering coefficient* for the jth scatterer and will be abbreviated to g_j.

The problem which we shall consider in this paper is then the following: Given the function $g(s, \omega)$ and the distribution function $n(\mathbf{r}, s)$ for the scatterers, and given also the form of the wave function $\psi_0(\mathbf{r})$ which is present in the medium in the absence of the scatterers (the incident wave), to find the configurational average of the wave function, $\psi(\mathbf{r})$, and of its mean-squared value, $|\psi(\mathbf{r})|^2$, in the presence of the scatterers.

FUNDAMENTAL EQUATIONS

Consider a particular configuration of the scatterers. We then attempt to represent the value of the wave function at the point $\mathbf{r}$ in the form

$$\psi(\mathbf{r})=\psi_0(\mathbf{r})+\sum_j A_jE(\mathbf{r}, \mathbf{r}_j),$$

where we have introduced the abbreviation

$$E(\mathbf{r}, \mathbf{r}')=\frac{\exp[-ik_0|\mathbf{r}-\mathbf{r}'|]}{|\mathbf{r}-\mathbf{r}'|}.$$

The equation represents the field as the sum of the incident wave and spherical waves diverging from each of the scatterers. The external field acting on the jth scatterer is then

$$\psi^j(\mathbf{r})=\psi_0(\mathbf{r})+\sum_{j'(\neq j)} A_{j'}E(\mathbf{r}, \mathbf{r}_{j'}).$$

We must then have, according to Eq. (6),

$$A_j=g_j\psi^j(\mathbf{r}_j),$$

so our equations become

$$\psi(\mathbf{r})=\psi_0(\mathbf{r})+\sum_j g_j\psi^j(\mathbf{r}_j)E(\mathbf{r}, \mathbf{r}_j), \quad (7)$$

$$\psi^j(\mathbf{r}_j)=\psi_0(\mathbf{r}_j)+\sum_{j'(\neq j)} g_{j'}\psi^{j'}(\mathbf{r}_{j'})E(\mathbf{r}_j, \mathbf{r}_{j'}). \quad (8)$$

These represent the fundamental equations of multiple scattering.

The direct method of solving our problem would then consist of solving the set of simultaneous linear algebraic Eqs. (8) for the $\psi^j(\mathbf{r}_j)$ and substituting these in Eq. (7), thus giving $\psi(\mathbf{r})$ as a function of the positions and scattering

parameters of the scatterers. Taking the configurational average of this quantity and of the square of its absolute value according to Eq. (3) would then give us the desired results. Unfortunately, it does not seem possible to carry this procedure through because of the complexity of the necessary integrations, and it is necessary to resort to another procedure. This alternative method consists of attempting to find equations satisfied by $\langle|\psi(\mathbf{r})|\rangle$ and $\langle|\psi(\mathbf{r})|^2\rangle$ and then solving these equations for the desired averaged quantities. We shall now carry through, with some approximations, this latter procedure.

THE CONFIGURATIONAL AVERAGE OF THE WAVE FUNCTION

Let us take the configurational average of both sides of Eq. (7) formally to obtain

$$\langle\psi(\mathbf{r})\rangle=\psi_0(\mathbf{r})+\sum_j\int_V\int g_j\langle\psi^j(\mathbf{r}_j)\rangle_j \times E(\mathbf{r},\mathbf{r}_j)\frac{n(\mathbf{r}_j,s_j)}{N}ds_j d\mathbf{r}_j$$

$$=\psi_0(\mathbf{r})+\int_V G(\mathbf{r}_j)\langle\psi^j(\mathbf{r}_j)\rangle_j E(\mathbf{r},\mathbf{r}_j)d\mathbf{r}_j, \qquad (9)$$

where

$$G(\mathbf{r})=\int g(s,\omega)n(\mathbf{r},s)ds. \qquad (10)$$

The quantity $\langle\psi^j(\mathbf{r}_j)\rangle_j$ represents the external field acting on the jth scatterer averaged over all possible configurations of all of the other scatterers. The only rigorous way of evaluating it seems to be to solve the set of simultaneous linear algebraic Eqs. (8) and then carry out the necessary integrations to obtain the above configurational average. This again is not feasible. We therefore must resort to the approximation of replacing the external field acting on the jth scatterer averaged over all configurations of the other scatterers by the average field which would exist at the position of the jth scatterer when the scatterer is not present. This last average field would differ only by a term of the order $1/N$ from[4] $\langle\psi(\mathbf{r}_j)\rangle$, so that if N is large, and if the above approximation is valid, which it appears to be on physical grounds, we may substitute $\langle\psi(\mathbf{r}_j)\rangle$ for $\langle\psi^j(\mathbf{r}_j)\rangle_j$ under the integral sign in the above equation thus obtaining the integral equation

$$\langle\psi(\mathbf{r})\rangle=\psi_0(\mathbf{r})+\int_V G(\mathbf{r}')\langle\psi(\mathbf{r}')\rangle E(\mathbf{r},\mathbf{r}')d\mathbf{r}', \qquad (11)$$

for the configurational average of the wave function.

To see the physical significance of this equation, let us apply the operator

$$\nabla^2+k_0^2 \qquad (12)$$

to both sides of the integral equation, remembering that

$$(\nabla^2+k_0^2)E(\mathbf{r},\mathbf{r}')=-4\pi\delta(\mathbf{r}-\mathbf{r}'), \qquad (13)$$

where $\delta(\mathbf{r}-\mathbf{r}')$ is the three-dimensional Dirac δ-function defined by the equation

$$\int_V f(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')d\mathbf{r}'=\begin{cases} f(\mathbf{r}) \text{ if } \mathbf{r} \text{ lies in } V \\ 0 \text{ if } \mathbf{r} \text{ lies outside of } V. \end{cases} \qquad (14)$$

We then obtain

$$\nabla^2\langle\psi(\mathbf{r})\rangle+k_0^2\langle\psi(\mathbf{r})\rangle=-4\pi G(\mathbf{r})\langle\psi(\mathbf{r})\rangle. \qquad (15)$$

If we write this as

$$\nabla^2\langle\psi(\mathbf{r})\rangle+k^2(r)\langle\psi(\mathbf{r})\rangle=0, \qquad (16)$$

where

$$k^2(\mathbf{r})=k_0^2+4\pi G(\mathbf{r}),$$

we see that $\langle\psi(\mathbf{r})\rangle$ satisfies the wave equation in a "continuous medium" in which the velocity of propagation depends upon the scattering coefficients and density of distribution of the scatterers and is, in general, a function of position. Thus the problem of finding the average value of the wave function has been essentially reduced to solving a boundary value problem in the wave equation. The boundary conditions are implied in the integral equation itself and depend on the function $G(\mathbf{r})$. If $G(\mathbf{r})$ is everywhere continuous and approaches a constant value or zero at infinity, then the boundary conditions are that $\langle\psi(\mathbf{r})\rangle-\psi_0(\mathbf{r})$ be everywhere continuous and have a continuous gradient and at infinity represent outward travelling waves. In another important case, that in which $G(\mathbf{r})$ is sectionally continuous, the boundary conditions are that $\langle\psi(\mathbf{r})\rangle-\psi_0(\mathbf{r})$ be everywhere continuous and have

[4] $\psi(\mathbf{r}_j)=\lim_{\mathbf{r}\to\mathbf{r}_j}\psi(\mathbf{r})$ is singular, but $\langle\psi(\mathbf{r}_j)\rangle=\lim_{\mathbf{r}\to\mathbf{r}_j}\langle\psi(\mathbf{r})\rangle$ is regular.

a continuous normal derivative across a surface of discontinuity of $G(\mathbf{r})$ and at infinity represent outward travelling waves, provided again that $G(\mathbf{r})$ approaches a constant value or zero at infinity. In both cases, of course $\langle\psi(\mathbf{r})\rangle$ must approach $\psi_0(\mathbf{r})$ as $G(\mathbf{r})$ becomes zero everywhere.

This reduction of the determination of $\langle\psi(\mathbf{r})\rangle$ to the solution of a boundary value problem in the wave equation corresponds, of course, to the well-known theory of the refractive index for light of gases.

We note also that the integral equation (11) can in principle be solved directly. One method of solution is the Liouville-Neumann method of successive substitutions or iteration method.[5] It consists in repeatedly substituting the expression for $\langle\psi(\mathbf{r})\rangle$ as given by the right-hand side of the integral equation for $\langle\psi(\mathbf{r}')\rangle$ under the integral sign, thus yielding in our case the infinite series

$$\langle\psi(\mathbf{r})\rangle=\sum_{m=0}^{\infty}\psi_m(\mathbf{r}) \tag{17}$$

where

$$\psi_m(\mathbf{r})=\int_V G(\mathbf{r}')\psi_{m-1}(\mathbf{r}')E(\mathbf{r},\mathbf{r}')d\mathbf{r}' \quad (m\neq 0) \tag{18}$$

which, if it converges uniformly, is the desired solution.

Reverting for a moment to the approximation $(\langle\psi^j(\mathbf{r}_j)\rangle_j\simeq\langle\psi(\mathbf{r}_j)\rangle)$ made at the beginning of this section, it should be mentioned that it would be desirable to demonstrate its validity mathematically rather than on intuitional grounds. Unfortunately it has not yet been found possible to do this. A similar approximation must be made in the next section to which this remark also applies.

THE CONFIGURATIONAL AVERAGE OF $|\psi(\mathbf{r})|^2$

To obtain the configurational average of $|\psi(\mathbf{r})|^2$ we follow an analogous procedure to that used for $\psi(\mathbf{r})$, but the analysis now becomes somewhat more involved. To begin we first multiply the expression for $\psi(\mathbf{r})$ as given by Eq. (7) by the corresponding expression for $\psi^*(\mathbf{r})$ with $\mathbf{r}$ set equal to $\mathbf{r}_0$, where the asterisk denotes the complex conjugate of the quantity to which it is attached. This yields

$$\begin{aligned}\psi(\mathbf{r})\psi^*(\mathbf{r}_0)&=\psi_0(\mathbf{r})\psi_0^*(\mathbf{r}_0)+\psi_0(\mathbf{r})[\psi^*(\mathbf{r}_0)-\psi_0^*(\mathbf{r}_0)]\\&+\psi_0^*(\mathbf{r}_0)[\psi(\mathbf{r})-\psi_0(\mathbf{r})]\\&+\sum_{j'}g_{j'}g_{j'}{}^*\psi^{j'}(\mathbf{r}_{j'})\psi^{j'*}(\mathbf{r}_{j'})E(\mathbf{r},\mathbf{r}_{j'})E^*(\mathbf{r}_0,\mathbf{r}_{j'})\\&+\sum_{j'j''}{}'g_{j'}g_{j''}{}^*\psi^{j'}(\mathbf{r}_{j'})\psi^{j''*}(\mathbf{r}_{j''})E(\mathbf{r},\mathbf{r}_{j'})E^*(\mathbf{r}_0,\mathbf{r}_{j''}).\end{aligned}$$

The prime on the last summation indicates that terms for which $j'=j''$ are to be omitted from the sum. If we now take the configurational average of both sides of this equation we obtain

$$\begin{aligned}\langle\psi(\mathbf{r})\psi^*(\mathbf{r}_0)\rangle&=\psi_0(\mathbf{r})\psi_0^*(\mathbf{r}_0)\\&+\psi_0(\mathbf{r})[\langle\psi^*(\mathbf{r}_0)\rangle-\psi_0^*(\mathbf{r}_0)]\\&+\psi_0^*(\mathbf{r}_0)[\langle\psi(\mathbf{r})\rangle-\psi_0(\mathbf{r})]\\&+\int_V H(\mathbf{r}_{j'})\langle\psi^{j'}(\mathbf{r}_{j'})\psi^{j'*}(\mathbf{r}_{j'})\rangle_{j'}\\&\quad\times E(\mathbf{r},\mathbf{r}_{j'})E^*(\mathbf{r}_0,\mathbf{r}_{j'})d\mathbf{r}_{j'}\\&+\frac{N-1}{N}\int_V\int_V G(\mathbf{r}_{j'})G^*(\mathbf{r}_{j''})\langle\psi^{j'}(\mathbf{r}_{j'})\psi^{j''*}(\mathbf{r}_{j''})\rangle_{j'j''}\\&\quad\times E(\mathbf{r},\mathbf{r}_{j'})E^*(\mathbf{r}_0,\mathbf{r}_{j''})d\mathbf{r}_{j'}d\mathbf{r}_{j''},\end{aligned} \tag{19}$$

where

$$H(\mathbf{r})=\int|g(s,\omega)|^2n(\mathbf{r},s)ds. \tag{20}$$

We now make a similar assumption to that made in the preceding section, although the physical meaning is not so clear in this case. We assume that it is valid to replace

$$\langle\psi^{j'}(\mathbf{r}_{j'})\psi^{j'*}(\mathbf{r}_{j'})\rangle_{j'} \text{ by } \langle\psi(\mathbf{r}_{j'})\psi^*(\mathbf{r}_{j'})\rangle$$

and

$$\langle\psi^{j'}(\mathbf{r}_{j'})\psi^{j''*}(\mathbf{r}_{j''})\rangle_{j'j''} \text{ by } \langle\psi(\mathbf{r}_{j'})\psi^*(\mathbf{r}_{j''})\rangle.$$

Making this substitution and replacing $(N-1)/N$ by unity since we are assuming N is large (compared to unity) we obtain the integral equation

$$\begin{aligned}\langle\psi(\mathbf{r})\psi^*(\mathbf{r}_0)\rangle&=\psi_0(\mathbf{r})\psi_0^*(\mathbf{r}_0)\\&+\psi_0(\mathbf{r})[\langle\psi^*(\mathbf{r}_0)\rangle-\psi_0^*(\mathbf{r}_0)]\\&+\psi_0^*(\mathbf{r}_0)[\langle\psi(\mathbf{r})\rangle-\psi_0(\mathbf{r})]\\&+\int_V H(\mathbf{r}')\langle\psi(\mathbf{r}')\psi^*(\mathbf{r}')\rangle E(\mathbf{r},\mathbf{r}')E^*(\mathbf{r}_0,\mathbf{r}')d\mathbf{r}'\\&+\int_V\int_V G(\mathbf{r}')G^*(\mathbf{r}'')\langle\psi(\mathbf{r}')\psi^*(\mathbf{r}'')\rangle\\&\quad\times E(\mathbf{r},\mathbf{r}')E^*(\mathbf{r}_0,\mathbf{r}'')d\mathbf{r}'d\mathbf{r}''.\end{aligned} \tag{21}$$

[5] Whittaker and Watson, *Modern Analysis* (Cambridge University Press, 1927), Chap. XI.

By substituting the right-hand side of Eq. (9) for $\langle\psi(\mathbf{r})\rangle$ and the complex conjugate for $\langle\psi^*(\mathbf{r}_0)\rangle$ in the above equation we obtain a simpler form of the equation:

$$\langle\psi(\mathbf{r})\psi^*(\mathbf{r}_0)\rangle-\langle\psi(\mathbf{r})\rangle\langle\psi^*(\mathbf{r}_0)\rangle$$
$$=\int_V H(\mathbf{r}')\langle\psi(\mathbf{r}')\psi^*(\mathbf{r}')\rangle E(\mathbf{r},\mathbf{r}')E^*(\mathbf{r}_0,\mathbf{r}')d\mathbf{r}'$$
$$+\int_V\int_V G(\mathbf{r}')G^*(\mathbf{r}'')[\langle\psi(\mathbf{r}')\psi^*(\mathbf{r}'')\rangle$$
$$-\langle\psi(\mathbf{r}')\rangle\langle\psi^*(\mathbf{r}'')\rangle]E(\mathbf{r},\mathbf{r}')E^*(\mathbf{r}_0,\mathbf{r}'')d\mathbf{r}'d\mathbf{r}''. \quad (22)$$

Now the above integral equation for $\langle\psi(\mathbf{r})\psi^*(\mathbf{r}_0)\rangle$ is in too complicated a form to obtain practical solutions of it directly. However, we can simplify the equation further in such a way that reasonably rapidly convergent series solutions may be obtained in many cases. An examination of the integral equation (22) shows that by obtaining the Liouville-Neumann iteration solution of the equation, considered as an integral equation in $[\langle\psi(\mathbf{r})\psi^*(\mathbf{r}_0)\rangle-\langle\psi(\mathbf{r})\rangle\langle\psi^*(\mathbf{r}_0)\rangle]$ with the first term on the right in Eq. (22) considered as the inhomogeneous term of the integral equation, $\langle\psi(\mathbf{r})\psi^*(\mathbf{r}_0)\rangle$ satisfies the integral equation

$$\langle\psi(\mathbf{r})\psi^*(\mathbf{r}_0)\rangle=\langle\psi(\mathbf{r})\rangle\langle\psi^*(\mathbf{r}_0)\rangle$$
$$+\int\langle\psi(\mathbf{r}')\psi^*(\mathbf{r}')\rangle L(\mathbf{r},\mathbf{r}_0;\mathbf{r}')d\mathbf{r}' \quad (23)$$

where the kernel $L(\mathbf{r},\mathbf{r}_0;\mathbf{r}')$ is given by the iteration series:

$$L(\mathbf{r},\mathbf{r}_0;\mathbf{r}')=\sum_{p=0}^{\infty}M_p(\mathbf{r},\mathbf{r}_0;\mathbf{r}'), \quad (24)$$

$$M_p(\mathbf{r},\mathbf{r}_0;\mathbf{r}')=\int_V\int_V G(\mathbf{r}'')G^*(\mathbf{r}''')$$
$$\times M_{p-1}(\mathbf{r}'',\mathbf{r}''';\mathbf{r}')E(\mathbf{r},\mathbf{r}'')$$
$$\times E^*(\mathbf{r}_0,\mathbf{r}''')d\mathbf{r}''d\mathbf{r}''', \quad (25)$$

$$M_0(\mathbf{r},\mathbf{r}_0;\mathbf{r}')=H(r')E(\mathbf{r},\mathbf{r}')E^*(\mathbf{r}_0,\mathbf{r}'). \quad (26)$$

This solution has the disadvantage that the series solution for the kernel $L(\mathbf{r},\mathbf{r}_0;\mathbf{r}')$ will be very slowly convergent since the factors $E(\mathbf{r},\mathbf{r}'')$ and $E^*(\mathbf{r}_0,\mathbf{r}''')$ in the iteration integrals are imaginary exponentials and do not promote convergence of the series, in general. An improvement on the situation is possible by obtaining a more rapidly convergent series for the kernel.

To do this we first find the partial differential equation satisfied by $L(\mathbf{r},\mathbf{r}_0;\mathbf{r}')$. If we apply the operator $(\nabla_0^2+k_0^2)(\nabla^2+k_0^2)$ to both sides of Eq. (24), where ∇^2 is the Laplacian operator in the $\mathbf{r}$ coordinates and ∇_0^2 the Laplacian operator in the $\mathbf{r}_0$ coordinates, the result of this operator acting on M_0 will be $16\pi^2\delta(\mathbf{r}-\mathbf{r}')\delta(\mathbf{r}_0-\mathbf{r}')$, while the result of the operator acting on the remainder of the series is readily found to be $16\pi^2G(\mathbf{r})\times G^*(\mathbf{r}_0)L(\mathbf{r},\mathbf{r}_0;\mathbf{r}')$. Thus the integral equation kernel satisfies the partial differential equation

$$[(\nabla_0^2+k_0^2)(\nabla^2+k_0^2)-16\pi^2G(\mathbf{r})G^*(\mathbf{r}_0)]L(\mathbf{r},\mathbf{r}_0;\mathbf{r}')$$
$$=16\pi^2\delta(\mathbf{r}-\mathbf{r}')\delta(\mathbf{r}_0-\mathbf{r}'). \quad (27)$$

If we now add

$$\{4\pi G(\mathbf{r})[\nabla_0^2+k^{*2}(\mathbf{r}_0)]$$
$$+4\pi G^*(\mathbf{r}_0)[\nabla^2+k^2(\mathbf{r})]\}L(\mathbf{r},\mathbf{r}_0;\mathbf{r}') \quad (28)$$

to both sides of this equation, and remember that

$$k^2(\mathbf{r})=k_0^2+4\pi G(\mathbf{r}), \quad (16)$$

and

$$k^{*2}(\mathbf{r}_0)=k_0^2+4\pi G^*(\mathbf{r}_0), \quad (16')$$

the equation can be written

$$[(\nabla_0^2+k^{*2}(\mathbf{r}_0)][\nabla^2+k^2(\mathbf{r})]L(\mathbf{r},\mathbf{r}_0;\mathbf{r}')$$
$$=16\pi^2\delta(\mathbf{r}-\mathbf{r}')\delta(\mathbf{r}_0-\mathbf{r}')$$
$$+\{4\pi G(\mathbf{r})[\nabla_0^2+k^{*2}(\mathbf{r}_0)]$$
$$+4\pi G^*(\mathbf{r}_0)[\nabla^2+k^2(\mathbf{r})]\}L(\mathbf{r},\mathbf{r}_0;\mathbf{r}'). \quad (29)$$

Let us now turn our attention to the function $K(\mathbf{r},\mathbf{r}')$ defined as the solution of the equation

$$\nabla^2K(\mathbf{r},\mathbf{r}')+k^2(\mathbf{r})K(\mathbf{r},\mathbf{r}')=-4\pi\delta(\mathbf{r}-\mathbf{r}'), \quad (30)$$

which represents outward travelling waves at infinity. It will be noted that $K(\mathbf{r},\mathbf{r}')$ bears the same relationship to Eq. (15) as $E(\mathbf{r},\mathbf{r}')$ does to Eq. (4). $K(\mathbf{r},\mathbf{r}')$ is a Green's function for Eq. (15) and can be interpreted as the wave field produced by a point source of unit strength placed at the point $\mathbf{r}'$ in a medium in which the propagation constant at the point $\mathbf{r}$ is given by $k(\mathbf{r})$. Then $K(\mathbf{r},\mathbf{r}')K^*(\mathbf{r}_0,\mathbf{r}')$ satisfies the equa-

tion

$$[\nabla_0^2+k^{*2}(\mathbf{r}_0)][\nabla^2+k^2(\mathbf{r})]K(\mathbf{r},\mathbf{r}')K^*(\mathbf{r}_0,\mathbf{r}') = 16\pi^2\delta(\mathbf{r}-\mathbf{r}')\delta(\mathbf{r}_0-\mathbf{r}'). \quad (31)$$

From this one readily finds that $L(\mathbf{r},\mathbf{r}_0;\mathbf{r}')$ satisfies the integral equation

$$L(\mathbf{r},\mathbf{r}_0;\mathbf{r}')=K(\mathbf{r},\mathbf{r}')K^*(\mathbf{r}_0,\mathbf{r}') + \frac{1}{4\pi}\int_V\int_V[\{G(\mathbf{r}'')[\nabla''^2+k^{*2}(\mathbf{r}'')] + G^*(\mathbf{r}''')[\nabla'''^2+k^2(\mathbf{r}''')]\}L(\mathbf{r}'',\mathbf{r}''';\mathbf{r}')] \times K(\mathbf{r},\mathbf{r}'')K^*(\mathbf{r}_0,\mathbf{r}''')d\mathbf{r}''d\mathbf{r}''', \quad (32)$$

as may be verified by substitution in Eq. (29). The Liouville-Neumann iteration procedure applied to this integral equation gives us a new series solution for the kernel of our integral equation:

$$L(\mathbf{r},\mathbf{r}_0;\mathbf{r}')=\sum_{p=0}^{\infty}L_p(\mathbf{r},\mathbf{r}_0;\mathbf{r}'), \quad (33)$$

$$L_p(\mathbf{r},\mathbf{r}_0;\mathbf{r}')=\frac{1}{4\pi}\int_V\int_V[\{G(\mathbf{r}'')[\nabla''^2+k^{*2}(\mathbf{r}''')] + G^*(\mathbf{r}''')[\nabla'''^2+k^2(\mathbf{r}'')]\} \times L_{p-1}(\mathbf{r}'',\mathbf{r}''';\mathbf{r}')]K(\mathbf{r},\mathbf{r}'') \times K^*(\mathbf{r}_0,\mathbf{r}''')d\mathbf{r}''d\mathbf{r}''', \quad (34)$$

$$L_0(\mathbf{r},\mathbf{r}_0;\mathbf{r}')=K(\mathbf{r},\mathbf{r}')K^*(\mathbf{r}_0,\mathbf{r}'). \quad (35)$$

This, in general, will be a quite rapidly converging series since $K(\mathbf{r},\mathbf{r}')$ contains a real negative exponential. The successive series terms will contain increasingly larger numbers of integrations involving the negative exponential and will thus die out rapidly. Thus, for example, the second term in the series is

$$L_1(\mathbf{r},\mathbf{r}_0;\mathbf{r}')=-\int_V G(\mathbf{r}'')K(\mathbf{r}'',\mathbf{r}') \times K(\mathbf{r},\mathbf{r}'')K^*(\mathbf{r}_0,\mathbf{r}')d\mathbf{r}'' - \int_V G^*(\mathbf{r}'')K^*(\mathbf{r}'',\mathbf{r}')K^*(\mathbf{r}_0,\mathbf{r}'')K(\mathbf{r},\mathbf{r}')d\mathbf{r}''. \quad (36)$$

The solution to our problem has thus been reduced to solving the integral equation

$$\langle|\psi(\mathbf{r})|^2\rangle=|\langle\psi(\mathbf{r})\rangle|^2 + \int_V H(\mathbf{r}')\langle|\psi(\mathbf{r}')|^2\rangle L(\mathbf{r},\mathbf{r};\mathbf{r}')d\mathbf{r}'. \quad (37)$$

In case the kernel $L(\mathbf{r},\mathbf{r};\mathbf{r}')$ can be accurately represented by the first term in the series (33), the integral equation becomes

$$\langle|\psi(\mathbf{r})|^2\rangle=|\langle\psi(\mathbf{r})\rangle|^2 + \int_V H(\mathbf{r}')\langle|\psi(\mathbf{r}')|^2\rangle|K(\mathbf{r},\mathbf{r}')|^2d\mathbf{r}'. \quad (38)$$

This latter equation is probably the most important practical result of our investigation.

AUXILIARY PHYSICAL QUANTITIES

There are several other quantities of importance in wave propagation theory and we shall now obtain expressions for the configurational average of these. The first of these is the gradient of the wave function. It is readily observed that the operation of taking the gradient and the integrations involved in obtaining the configurational average commute so that the average of the gradient of the wave function is simply the gradient of the configurational average of the wave function:

$$\langle\nabla\psi(\mathbf{r})\rangle=\nabla\langle\psi(\mathbf{r})\rangle. \quad (39)$$

The other quantity of importance is the flux which is proportional to the quantity

$$\mathfrak{F}=\psi^*\nabla\psi-\psi\nabla\psi^*. \quad (40)$$

We may obtain the configurational average of this quantity by taking the gradient of both sides of Eq. (23) with respect to $\mathbf{r}$ and subtracting from this the gradient of both sides of the same equation with respect to $\mathbf{r}_0$. On subsequently setting $\mathbf{r}_0=\mathbf{r}$, we obtain

$$\langle\mathfrak{F}(\mathbf{r})\rangle=\langle\psi^*(\mathbf{r})\nabla\psi(\mathbf{r})-\psi(\mathbf{r})\nabla\psi^*(\mathbf{r})\rangle =\langle\psi^*(\mathbf{r})\rangle\nabla\langle\psi(\mathbf{r})\rangle-\langle\psi(\mathbf{r})\rangle\nabla\langle\psi^*(\mathbf{r})\rangle + \int_V H(\mathbf{r}')\langle|\psi(\mathbf{r}')|^2\rangle[\nabla L(\mathbf{r},\mathbf{r}_0;\mathbf{r}') - \nabla_0 L^*(\mathbf{r},\mathbf{r}_0;\mathbf{r}')]_{\mathbf{r}_0=\mathbf{r}}d\mathbf{r}'. \quad (41)$$

Again if $L(\mathbf{r}, \mathbf{r}_0; \mathbf{r}')$ can be replaced by the first term in its iteration series, this result can be written

$$\langle \mathfrak{F}(\mathbf{r}) \rangle = \langle \psi^*(\mathbf{r}) \rangle \nabla \langle \psi(\mathbf{r}) \rangle - \langle \psi(\mathbf{r}) \rangle \nabla \langle \psi^*(\mathbf{r}) \rangle + \int H(\mathbf{r}') \langle |\psi(\mathbf{r}')|^2 \rangle [K^*(\mathbf{r}, \mathbf{r}') \nabla K(\mathbf{r}, \mathbf{r}') - K(\mathbf{r}, \mathbf{r}') \nabla K^*(\mathbf{r}, \mathbf{r}')] d\mathbf{r}'. \quad (42)$$

Thus $\langle \mathfrak{F} \rangle$ can be calculated when $\langle |\psi(\mathbf{r})|^2 \rangle$ is known.

PHYSICAL INTERPRETATION

In order to interpret the results which we have obtained, it is useful first to consider the scattering of waves by a single scatterer. In this case the wave function becomes simply

$$\psi(\mathbf{r}) = \psi_0(\mathbf{r}) + g_1 \psi_0(\mathbf{r}_1) E(\mathbf{r}, \mathbf{r}_1), \quad (43)$$

which is the sum of the incident wave and the spherical wave scattered from the scatterer. If we consider a plane incident wave

$$\psi_0(\mathbf{r}) = A \exp[-i\mathbf{k}_0 \cdot \mathbf{r}], \quad (44)$$

and integrate the flux per unit area in the scattered wave, $\mathfrak{F} = \psi_s^* \nabla \psi_s - \psi_s \nabla \psi_s^*$, where ψ_s represents the second term on the right in Eq. (43), over a sphere whose center is at the position of the scatterer, we readily find that this total flux is

$$4\pi i k_0 |g_1|^2 A^2. \quad (45)$$

If we divide by the flux per unit area normal to $\mathbf{k}_0$ in the incident wave, $ik_0 A^2$, we obtain what is usually referred to as the *scattering cross section*, σ_s:

$$\sigma_s = 4\pi |g_1|^2. \quad (46)$$

It represents an area which when multiplied by the incident flux per unit area, gives the total flux in the scattered wave.

Now if one integrates the flux per unit area caused by both incident and scattered waves [that is $\psi^* \nabla \psi - \psi \nabla \psi^*$, where ψ is given by (43)] over the same sphere, one obtains

$$4\pi i k_0 |g_1|^2 A^2 + 4\pi (g_1 - g_1^*) A^2. \quad (47)$$

The negative of this quantity represents the net inward flux through the sphere. Dividing the latter by the flux per unit area in the incident wave gives the so-called *absorption* or *capture cross section*, σ_c, since it represents the area which when multiplied by the incident flux per unit area normal to $\mathbf{k}_0$, gives the flux which is effectively lost from the wave field. Thus

$$\sigma_c = -4\pi |g_1|^2 - \frac{4\pi Im(g_1)}{k_0} \equiv -\sigma_s + \sigma_e, \quad (48)$$

where

$$\sigma_e = -4\pi Im(g_1)/k_0 \quad (49)$$

is usually referred to as the *extinction cross section*, being the sum of the capture and scattering cross sections. In practical problems, the imaginary part of g_1 is usually negative so that the extinction cross section is positive.[6] Expressed in terms of the more familiar cross sections, the scattering coefficient g_1 becomes

$$g_1 = \left(\frac{\sigma_s}{4\pi} - \frac{k_0^2 \sigma_e^2}{16\pi^2} \right)^{\frac{1}{2}} - i \frac{k_0 \sigma_e}{4\pi}. \quad (50)$$

We may then define a scattering cross section per unit volume, $S_s(\mathbf{r})$, by

$$S_s(\mathbf{r}) = \int \sigma_s(s) n(\mathbf{r}, s) ds = 4\pi \int |g(s)|^2 n(\mathbf{r}, s) ds = 4\pi H(\mathbf{r}), \quad (51)$$

and an extinction cross section per unit volume, $S_e(\mathbf{r})$, by

$$S_e(\mathbf{r}) = \int \sigma_e(s) n(\mathbf{r}, s) ds = -\frac{4\pi}{k_0} \int Im[g(s)] n(\mathbf{r}, s) ds = -\frac{4\pi}{k_0} Im[G(\mathbf{r})]. \quad (52)$$

Let us now examine Eq. (15) which governs the propagation of the average value of the wave function in the case where $G(\mathbf{r})$ is a constant independent of position. In this case the plane wave solution of the differential equation will be

[6] If there is no "true" capture, σ_c must be zero. For a particular field, this can be shown from the field equations and, where appropriate, the boundary conditions at the scatterer. However, g will still be complex because of "radiation resistance."

of the form

$$\langle\psi(\mathbf{r})\rangle = A \exp[-i\mathbf{k}\cdot\mathbf{r}], \tag{53}$$

where $|\mathbf{k}| = k$ and the direction of $\mathbf{k}$ is the direction of propagation of the wave. Now since k is in general complex the wave represented in Eq. (53) is attenuated with distance. Now we have

$$k^2 = k_0^2 + 4\pi G = k_0^2 + 4\pi Re(G) + 4\pi i Im(G), \tag{54}$$

so that when the number of scatterers per unit volume is sufficiently small we may expand the square root for k to obtain

$$k \simeq k_0 + \frac{2\pi}{k_0} Re(G) + \frac{2\pi i}{k_0} Im(G). \tag{55}$$

Thus the flux in the wave (53) is reduced by the factor

$$\left[\exp \frac{2\pi}{k_0} Im(G)\right]^2 = e^{-S_e} \tag{56}$$

per unit distance. Hence in this case the medium acts as if it removes in a unit distance a fraction equal to the extinction cross section per unit volume of the flux carried by the wave. Alternatively, we may say that the medium acts as if each scatterer removes from the wave an amount of flux equal to its extinction cross section times the flux per unit area incident on it in the average wave. This flux is partially absorbed (if the capture cross section is not zero) and partially scattered out of the average wave propagating through the medium. It should be noted that this simple picture is applicable only when the number of scatterers per unit volume is sufficiently small so that the above expansion of the square root is valid. Otherwise there is interference between the scatterers and they no longer can be considered to absorb and scatter independently.

In addition the phase velocity $c(\mathbf{r})$ of the average wave is different from the phase velocity in the scatterer-free medium:

$$\frac{1}{c^2} = \frac{1}{c_0^2} + \frac{4\pi G(\mathbf{r})}{\omega^2}. \tag{57}$$

This may be attributed to the fact that there is a difference in phase between the wave incident on a scatterer and the scattered wave from the scatterer. When the scattered waves are combined with the incident wave there is a progressive advancement or retardation of the phase of the disturbance and the result appears as a change in the phase velocity.[7]

Let us now turn our attention to the physical interpretation of our integral equation for the configurational average of the square of the wave function written in terms of the scattering cross section per unit volume [see Eqs. (51), (37)]:

$$\langle|\psi(\mathbf{r})|^2\rangle = |\langle\psi(\mathbf{r})\rangle|^2 + \frac{1}{4\pi}\int S_s(\mathbf{r}')\langle|\psi(\mathbf{r}')|^2\rangle L(\mathbf{r}, \mathbf{r}; \mathbf{r}')d\mathbf{r}'. \tag{58}$$

Assume first that we may replace $L(\mathbf{r}, \mathbf{r}; \mathbf{r}')$ by the first term in the series (33), $|K(\mathbf{r}, \mathbf{r}')|^2$. Now we have already pointed out that $K(\mathbf{r}, \mathbf{r}')$ can be interpreted as the value of the wave function at the point $\mathbf{r}$ caused by a unit point source at the point $\mathbf{r}'$, in the continuous medium [characterized by the propagation constant $k(\mathbf{r})$] in which $\langle\psi(\mathbf{r})\rangle$ is propagated. Thus in this case the integral equation above simply states that the value of $\langle|\psi(\mathbf{r})|^2\rangle$ at any point $\mathbf{r}$ is equal to the mean-square value of the average wave at the point $\mathbf{r}$, $|\langle\psi(\mathbf{r})\rangle|^2$, plus contributions owing to scattered waves from all other points of the medium, the strength of the sources per unit volume at these other points being proportional to the value of $\langle|\psi(\mathbf{r})|^2\rangle$ at these points and to the scattering cross section per unit volume at these points. The various contributions, it will be noticed, combine as if the waves to which they are owing are *incoherent*, that is, bear a random phase relationship to one another. If we obtain the series solution of this integral equation by the Liouville-Neumann method, this solution, too, has a simple physical interpretation. The solution can be written

$$\langle|\psi(\mathbf{r})|^2\rangle = \sum_{m=0}^{\infty} \Psi_m(\mathbf{r}), \tag{59}$$

$$\Psi_m(\mathbf{r}) = \frac{1}{4\pi}\int S_s(\mathbf{r}')\Psi_{m-1}(\mathbf{r}')|K(\mathbf{r}, \mathbf{r}')|^2 d\mathbf{r}', \tag{60}$$

$$\Psi_0(\mathbf{r}) = |\langle\psi(\mathbf{r})\rangle|^2. \tag{61}$$

[7] The condition for no dispersion in the presence of the scatterers is obviously: $g(\omega, s) = \omega^2\gamma(s)$, where γ is independent of ω.

Thus the first term in the series represents the contribution caused by the average wave, the second term, the contributions caused by a single scattering of the average wave from each point in the medium to the point in question, the third term, the contributions caused by scattering twice of the average wave, etc. Hence, when $L(\mathbf{r}, \mathbf{r}; \mathbf{r}')$ can be replaced by the first term in its series solution, we have a very simple and logical physical interpretation of the integral equation which we have derived.

When the full series solution for $L(\mathbf{r}, \mathbf{r}; \mathbf{r}')$ is substituted in the integral equation, however, the meaning of the additional terms is not so clear. An examination of them indicates that they are some sort of residual interference terms which are not eliminated in the averaging process. This is apparent from the fact that the imaginary exponential terms are still present [Eqs. (34), (36)]. The physical interpretation of each term is quite simple: Thus, the terms corresponding to the next term in the series for L, that is L_1 [Eq. (36)], may be interpreted as the resultant at the point $\mathbf{r}$ of a wave scattered from the point $\mathbf{r}'$ to $\mathbf{r}$ by two paths, a direct path from $\mathbf{r}'$ to $\mathbf{r}$ and also by propagation from $\mathbf{r}'$ to another point $\mathbf{r}''$ from which the wave is rescattered *coherently* to the point $\mathbf{r}$, the result being summed over all intermediate points $\mathbf{r}''$. Similarly the higher terms of the series correspond to larger numbers of intermediate coherent scatterings. It is difficult to see, however, why these additional terms should be present in the equation. A rough calculation of the order of magnitude of these terms shows that in many practical cases, they are not very significant. Thus the ratio of the terms caused by L_1 to the terms caused by $L_0 = |K(\mathbf{r}, \mathbf{r}')|^2$ seems to be of the order of magnitude of $(k^2 - k_0^2)/k_0^2$, so that if the change in the velocity of the waves in the medium produced by the introduction of the scatterers is small, these higher terms are of little importance.

In conclusion, it may be helpful to point out that in most measurements on waves it is the mean-square value of the wave function rather than the wave function itself that is measured; therefore $\langle|\psi(\mathbf{r})|^2\rangle$ will usually be of more interest than $\langle\psi(\mathbf{r})\rangle$. Now the difference between $\langle|\psi(\mathbf{r})|^2\rangle$ and $|\langle\psi(\mathbf{r})\rangle|^2$ can be interpreted from the view of our statistical ensemble of configurations as effected by the statistical fluctuations of $\psi(\mathbf{r})$ about $\langle\psi(\mathbf{r})\rangle$ in different members of the ensemble. Another method of viewing this difference may be helpful in bringing out the physical interpretation. Consider instead of a statistical ensemble of configurations of scatterers, a collection of scatterers which are moving about slowly (so that Doppler effects may be neglected) in such a way that over a period of time the collection passes through all the configurations of our original ensemble. In this case $\langle|\psi(\mathbf{r})|^2\rangle$ can be interpreted as the long time average of $|\psi(\mathbf{r}, t)|^2$, $\psi(\mathbf{r}, t)$ being the value of the wave function at the point $\mathbf{r}$ at the time t. For such a collection of scatterers $\langle\psi(\mathbf{r})\rangle$ has little physical significance unless

$$\langle|\psi(\mathbf{r})|^2\rangle - |\langle\psi(\mathbf{r})\rangle|^2 \ll \langle|\psi(\mathbf{r})|^2\rangle. \tag{62}$$

To define the meaning of $\langle\psi(\mathbf{r})\rangle$ precisely, one may say that if $\psi(\mathbf{r}, t)$ represents the value of the wave function at the point $\mathbf{r}$ at time t, then $\langle\psi(\mathbf{r})\rangle$ is the function which makes the integral

$$\int |\psi(\mathbf{r}, t) - \langle\psi(\mathbf{r})\rangle e^{i\omega t}|^2 dt \tag{63}$$

a minimum, the integration being taken over a length of time sufficient for the collection of scatterers to have passed through all the configurations of the ensemble.

On the other hand, we have still to indicate why in scattering by a fixed configuration of scatterers, "incoherent scattering" is observed. The answer to this question lies in the fact that the regularly propagated wave or "coherent scattering" is precisely defined only as an average over the ensemble of configurations. For a fixed configuration, all of the scattering is strictly coherent but is artificially divided into a part which is estimated to propagate uniformly and a remainder which is referred to as "incoherently scattered." The basis of this distinction lies in the fact that for a collection of a large number of scatterers a particular "unprepared" collection will have particular physical properties which do not deviate greatly from the average physical properties of a properly defined statistical ensemble of collections because of the lack of "correlation" implied in the word "unprepared" as

to the positions of the individual scatterers. This allows one to estimate the properties of the average over an ensemble from an experiment on a particular configuration with a very high probability that the estimate is correct, although the possibility exists of a wide deviation if the selected ensemble should be particularly strongly ordered. In experiments usually several samples are used so as to guard against this remote possibility. We cannot enter further into this discussion of the relationship between statistical concepts and experiment, but refer the reader to almost any text on statistical mechanics for further elucidation.

DISCUSSION

It is of interest to indicate the directions in which the present theory may be generalized to embrace a wider range of phenomena:

(1) We may attempt to lift the restriction to isotropic (*s*-) scattering. This could be accomplished by including higher order scattering terms. Thus, in the next approximation we would include terms corresponding to dipole (*p*-) scattering which are of the form

$$\mathbf{B}\cdot\nabla E(\mathbf{r}, \mathbf{r}_j), \tag{64}$$

where **B** is some vector function of the external field acting on the scatterer, such as its gradient. By adding further terms corresponding to quadripole scattering, etc., one could, in principle, adequately treat any type of scattered wave. The difficulties in this extension of the theory appear to be simply ones of increasing complexity of the expressions handled.

(2) One might attempt to consider more general forms of probability distribution functions than those included in Eq. (1). Such more general functions would involve correlations between the positions of various scatterers so that the probability of finding a particular scatterer in the neighborhood of some point would be a function of the positions of one or more of the other scatterers. This type of function would allow a more accurate representation of the state of affairs in "amorphous" solids, liquids, and imperfect gases where such correlations do exist. Tempting as such a generalization might be, the well-known difficulties which arise in averaging over correlative probability distribution functions, are discouraging of elegant attempts at a rigorous theory in this direction. Some promise might appear, however, in an approximate treatment of this problem by simplifying the correlation between scatterers to that between the position of each scatterer and the average distribution of the remainder such as is done in the well-known approximate treatments of cooperative phenomena in statistical mechanics, or by the use of such methods as are employed in the theory of x-ray (single) scattering in liquids.

(3) A further generalization may be made in the extension of the above treatment to vector waves. If the vector wave function satisfies the vector wave equation and if the wave scattered by a scatterer is proportional directly to the external field acting on the scatterer, then the generalization follows immediately. However, in the case of greatest practical importance, electromagnetic waves, the above is not true. The wave scattered by a scatterer is to first order proportional, not to the external field acting on the scatterer, but to the curlcurl of this field.[8] This introduces singularities into the integrands of the integral equations which must be omitted from the integration by enclosing them by small spheres. These singularities lead in the treatment of the average value of the wave function to the well-known Lorentz-Lorenz effect,[3] but the derivation along rigorous lines is long and involved. These singularities appear to cause still more difficulty in the treatment of the average value of the square of the wave function (electric or magnetic field strength) but there seems no reason why the analysis cannot be carried through. The problem is of sufficient practical importance as well as theoretical importance to warrant additional study.

(5) Another generalization could be attempted in the direction of considering changes of frequency of scattered waves produced through the Doppler effect by moving scatterers. This is closely related to the problem of inelastic scat-

[8] The vector wave function is the Hertz vector **Z** in this case; the scattered waves are of the form $\mathbf{p}E(\mathbf{r}, \mathbf{r}_j)$, where **p**, the dipole moment of the scatterer is proportional to the external electric field acting on the scatterer, which is in turn proportional to the curlcurl of the external **Z** field acting on the scatterer.

tering of light and fundamental particles by free or bound atoms in the quantum theory, since a change of frequency of the wave function corresponds in quantum mechanics to a change in energy of the associated particle. Such a theory would be developed along lines in some ways similar to the treatment of diffuse spots in the theory of x-ray diffraction which are also the effect of inelastic scattering of photons by the crystal lattice. Such a theory might have considerable value in investigating experimentally the dynamics of thermal agitation in amorphous solids, liquids, and gases.

(6) Finally in a similar manner, the theory could be extended to include non-steady state scattering, in which the incident wave function does not have a harmonic dependence on the time.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Dr. H. Primakoff for many helpful discussions on this subject, to Dr. R. S. Shankland for his continual encouragement, and to Dr. E. Gerjuoy for bringing to the author's attention a significant error in an earlier draft of this paper.

Letters to the Editor

PROMPT publication of brief reports of important discoveries in physics may be secured by addressing them to this department. The closing date for this department is the third of the month. Because of the late closing date for the section no proof can be shown to authors. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents. Communications should not in general exceed 600 words in length.

Double Bragg Reflections of X-Rays in a Single Crystal

C. J. DAVISSON AND F. E. HAWORTH
Bell Telephone Laboratories, Murray Hill, New Jersey
January 25, 1945

IT has been called to our attention that experiments similar to those which we recently described under the above title in these columns,[1] though made on a different substance, were made and published eight years ago by M. Renninger in Germany.

It is a matter of some embarrassment that we were unacquainted with Renninger's work, and of regret that we failed, in what was obviously an incomplete search of the literature, to find the articles in which it is described.

The phenomenon which we observed in quartz and designated as "double Bragg reflection" is identical with that which Renninger[2] observed in diamond and designated by the somewhat inappropriate name of *Umweganregung*, which may be translated as "circuitous excitation" or "indirect excitation."

[1] C. J. Davisson and F. E. Haworth, Phys. Rev. 66, 351 (1944).
[2] Renninger, Zeits. f. Physik 106, 141 (1937); Zeits. f. Krist. 97, 107 (1937).

The Red-Giant Stage of Stellar Evolution

G. GAMOW
The George Washington University, Washington, D. C.
January 24, 1945

WHEN a normally evolving main-sequence star (point-source model with the carbon-cycle energy production) consumes all hydrogen contained in its central regions, one must expect the reaction zone to move outwards thus leading to the so-called "shell-source" model[1] consisting of the isothermal dehydrogenized core and the radiative envelope.

It was previously suggested by the author[2] that the progressive growth of the energy-producing shell must result in a large increase of stellar radius, thus bringing the star into the red-giant region of the Hertzsprung-Russell diagram. This point of view encountered, however, serious difficulties owing to the fact that the exact mathematical treatment of the isothermal core model[3] seemed to lead to the conclusion that no such model is possible if more than ten percent of total stellar mass is to be contained in the core.

It must be indicated here that the above paradoxical result is entirely owing to the arbitrary assumption that the isothermal core of the star must be governed by the ideal gas law, and disappears at once if we take into account the possibility of degeneracy near the center of the core. Indeed the impossibility of fitting in a massive isothermal core of ideal gas (polytrope $n=\infty$) rests on the fact that above a certain density limit, the gas pressure in the center of the core is unable to support the weight of the core thus leading to an unlimited contraction. Physically, however, such a contraction will be rapidly stopped because of the decreased compressibility of gas in the degenerated region. The stable state of a star with a core exceeding ten percent of the total mass, will thus consist of three different regions: (1) degenerated nucleus, (2) isothermal layer of ideal gas, and (3) radiative envelope. The size and the luminosity of the star will be completely determined by the condition of degeneracy at the boundary between (1) and (2), and the condition of energy-production (essentially $T \cong 2 \cdot 10^7$ °K) at the boundary between (2) and (3). Preliminary calculations of such composite stellar model, based on fitting of partially degenerated isothermal solutions for the core to the centrally condensed solution for the envelope, indicate that the growth of energy-producing shell may increase the radius of the star by a factor of few hundred, and lower its surface temperature by a factor two or three.

Since more massive stars develop the shell at a much higher rate, the locus of shell-source stars in the Hertzsprung-Russell diagram must be represented by a line branching off from the main-sequence in the direction of higher luminosities and lower surface temperatures. Assuming the age of the main bulk of stars to be about 3×10^9 years, we find that this branching off must take place near $+2$ absolute magnitude. It seems very likely that the red-giant branch of the stellar population II (elliptic nebulae, nuclei of spiral nebulae, and globular clusters) recently discussed by Baade,[4] corresponds to the group of (pregalactic) stars passing through this particular stage of their evolution. This would lead us to a very interesting conclusion that the members of the stellar population I (spiral arms) are cosmologically considerably younger than those of the population II.

It may be also expected that with the gradual approach of the energy-producing shell to the stellar surface, the star will begin to eject its outer layers into the surrounding space, thus passing rapidly into the so-called Wolf-Rayet stage of evolution. The resulting loss of a large part of stellar mass must lead finally to the unveiling of the white dwarf, which was gradually growing in the center of the star during the red-giant part of its evolution. This view may find some support in the fact that the known white dwarf stars, which are apparently deprived of any nuclear energy-source, possess internal temperatures of the order 2×10^7 °K which corresponds to the carbon-cycle source. This could be indeed interpreted by saying that those dense, slowly-cooling bodies still retain the temperature of

the isothermal core of the red-giant stars in which they must have been formed.

The results of more detailed calculations of the above "three-layer" stellar model, and also the discussion of possible astrophysical consequences will be published in due course.

[1] G. Gamow, Phys. Rev. 53, 595 (1938); Chas. Critchfield and G. Gamow, Astrophys. J. 89, 244 (1939).
[2] G. Gamow, Phys. Rev. 65, 20 (1944).
[3] S. Chandrasekhar and L. R. Henrich, Astrophys. J. 94, 525 (1942); S. Chandrasekhar and M. Schoenberg, Astrophys. J. 96, 161 (1942).
[4] W. Baade, Astrophys. J. 100, 137 (1944).

A Note on the Kepler Problem in a Space of Constant Negative Curvature

L. INFELD AND A. SCHILD
Department of Mathematics, University of Toronto, Toronto, Canada
January 10, 1945

SCHRÖDINGER[1] recently solved the "Kepler problem" in a spherical or Einstein universe and obtained the interesting result that the energy spectrum is discrete everywhere. It is instructive to compare this spectrum with that in an "open" universe of constant negative curvature, which is in fact Milne's universe.[2] As will be shown, the spectrum consists of a *finite* number of (mostly negative) energy levels in addition to a continuous spectrum. To our knowledge, this is the first quantum mechanical problem to exhibit a finite number of discrete energy values.

The line element of the hyperbolic 3-space may be written in the form:

$$ds^2=R^2d\alpha^2+R^2\sinh^2\alpha(d\theta^2+\sin^2\theta d\phi^2). \tag{1}$$

Letting $R\to\infty$, $\alpha\to 0$ such that $\alpha R=r$ remains finite, this line element reduces to that of Euclidean space, in which r, θ, ϕ are the usual polar coordinates. Schrödinger's equation is taken in the form

$$\Delta\psi+(2\mu/\hbar^2)(E-V)\psi=0, \tag{2}$$

with the proper Laplacian belonging to (1).

The potential energy $V(\alpha)$ of an electron in a central Coulomb field must satisfy Laplace's equation

$$\Delta V=g^{ij}V_{|ij}=0.$$

We find

$$V=-(Ze^2/R)(\coth\alpha-1), \tag{3}$$

if we demand that $V\to 0$ as $\alpha\to\infty$ and $V\to -Ze^2/r$ as $R\to\infty$, $\alpha R\to r$.

Putting $\psi=\sigma(\alpha)Y_l(\theta,\phi)$ in Eq. (2), where the Y_l are normalized spherical harmonics, we obtain the radial equation

$$(d/d\alpha)(\sinh^2\alpha d\sigma/d\alpha)+(\lambda+1-2\nu)\sinh^2\alpha\,\sigma+2\nu\sinh\alpha\cosh\alpha\,\sigma-l(l+1)\sigma=0; \tag{4}$$

where $\lambda=2\mu R^2E/\hbar^2-1$ and $\nu=Ze^2\mu R/\hbar^2$.

Introducing the "density function" $\rho=\sigma\sinh\alpha$, we can factorize Eq. (4):

$$\{(l+1)\coth\alpha-\nu/(l+1)+d/d\alpha\}\times\{(l+1)\coth\alpha-\nu/(l+1)-d/d\alpha\}\rho=\{\lambda-2\nu+(l+1)^2+\nu^2/(l+1)^2\}\rho; \tag{5}$$

$$\{l\coth\alpha-\nu/l-d/d\alpha\}\{l\coth\alpha-\nu/l+d/d\alpha\}\rho=\{\lambda-2\nu+l^2+\nu^2/l^2\}\rho. \tag{5'}$$

The only possible discrete solutions are immediately obtained by the factorization method. We have, in the notation of Infeld's paper,[1]

$$\lambda_n=2\nu-(n+1)^2-\nu^2/(n+1)^2;\quad n=0,1,2\cdots, \tag{6}$$

$$\rho_n{}^n=c\sinh^{n+1}\alpha\exp(-\nu\alpha/(n+1)); \tag{7}$$

$$\rho_n{}^{l-1}=\{l^2+\nu^2/l^2-(n+1)^2-\nu^2/(n+1)^2\}^{-\frac{1}{2}}\times(l\coth\alpha-\nu/l+d/d\alpha)\rho_n{}^l;\quad l=0,1\cdots n, \tag{7'}$$

since $\sinh\alpha\sim\frac{1}{2}e^\alpha$ for large α, Eq. (7) shows that if the normalization integral for $\rho_n{}^n$ is to exist, we must have

$$(n+1)^2<\nu;\quad n=0,1\cdots n_0, \tag{8}$$

where n_0 is the greatest integer satisfying the inequality.

Thus we have a finite number of discrete energy levels:

$$E_n=Ze^2/R-\hbar^2n(n+2)/2\mu R^2-Z^2e^4\mu/2\hbar^2(n+1)^2;\quad n=0,1\cdots n_0. \tag{9}$$

It is interesting to note that the number of discrete levels $n_0+1\sim\nu^{\frac{1}{2}}=(R/a)^{\frac{1}{2}}$, where a is the radius of the first Bohr orbit of the hydrogen-like atom. Taking R to be of the order of 10^{28} cm (the usual magnitude of the radius of the universe considered in relativistic cosmology), we find that n_0 is a large number of order 10^{18}. The highest discrete energy level E_{n_0} lies between $-3\hbar^2/2\mu R^2$ and $\hbar^2/2\mu R^2$ and may thus be positive or negative; all others are negative.

The factorization method breaks down for the continuous spectrum. Putting $x=\coth\alpha$, Eq. (4) transforms into

$$\frac{d^2\sigma}{dx^2}+\left[\frac{\lambda+1-2\nu+2\nu x}{(x^2-1)^2}-\frac{l(l+1)}{x^2-1}\right]\sigma=0. \tag{10}$$

The range of x (for $0\leqslant\alpha<\infty$) is from ∞ to 1. The only solution of this Riemann P-equation, bounded for large x, is[3]

$$\sigma=(x+1)^{-l-\frac{1}{2}(1+\sqrt{-\lambda})}(x-1)^{\frac{1}{2}(1+\sqrt{-\lambda})}\times F(1+l+\tfrac{1}{2}[-\lambda^{\frac{1}{2}}-(4\nu-\lambda)^{\frac{1}{2}}],\ 1+l+\tfrac{1}{2}[-\lambda^{\frac{1}{2}}+(4\nu-\lambda)^{\frac{1}{2}}];\ 2(l+1);\ 2/(x+1)). \tag{11}$$

This solution is continuous everywhere in our interval (including ∞), except possibly at $x=1$. For negative λ we regain the discrete spectrum discussed above.

If λ is positive, $-\lambda^{\frac{1}{2}}=i\lambda^{\frac{1}{2}}$ is imaginary. Introducing α for x in (11) and replacing σ by the "density function" ρ, we have

$$\rho=\sinh^{l+1}\alpha\exp(-(1+i\lambda^{\frac{1}{2}}+l)\alpha)\times F(1+l+\tfrac{1}{2}[i\lambda^{\frac{1}{2}}-(4\nu-\lambda)^{\frac{1}{2}}],\ 1+l+\tfrac{1}{2}[i\lambda^{\frac{1}{2}}+(4-\nu-\lambda)^{\frac{1}{2}}];\ 2(l+1);\ z);\quad z=2e^{-\alpha}\sinh\alpha. \tag{12}$$

We demand that ρ be continuous and bounded for $0\leqslant\alpha<\infty$. We have seen that ρ is continuous, and it is clear that the factor multiplying the hypergeometric function in (12) remains bounded for large α. It remains to be shown that the hypergeometric function is also bounded as $\alpha\to\infty$, i.e., as $z\to 1$.

Putting $z=1$, we find the ratio of successive terms of the hypergeometric series is of the form[4]

$$a_n/a_{n+1}=1+(1-i\lambda^{\frac{1}{2}})/n+0(n^{-2}).$$

It follows[5] that the infinite hypergeometric series oscillates finitely. Abel's theorem[6] then shows that the hypergeometric function in (12) remains bounded as $z \to 1$ or $\alpha \to \infty$.

This establishes the existence of a continuous eigenvalue spectrum of all positive λ or

$$E > \hbar^2/2\mu R^2. \tag{13}$$

In order to obtain an indication of how a relativistic treatment might affect this problem, the authors repeated the calculations by use of the Gordon-Klein equation. The general character of the solution was found to be unchanged.

The authors wish to thank Professor A. F. Stevenson for helpful discussion.

[1] E. Schrödinger, Proc. Roy. Irish Acad. **A46**, 9 (1940). Also, L. Infeld, Phys. Rev. **59**, 737 (1941) and A. F. Stevenson, Phys. Rev. **59**, 842 (1941).
[2] C. Gilbert, Quart. J. Math. **9**, 187, Eq. (9) (1938), and A. G. Walker, M. N. R. A. S. **95**, 263 (1935).
[3] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge, 1940), Secs. 10.7, 10.71, 10.72 and Chaps. 14.
[4] T. J. I. Bromwich, *An Introduction to the Theory of Infinite Series* (The Macmillan Company, New York, 1926), Sec. 79, p. 241, example.
[5] Reference 4, Sec. 79, p. 241, rule (ii).
[6] Reference 4, Sec. 51, p. 149.

Classical Theory of the Point Electron

MARIO SCHÖNBERG
Department of Physics, University of São Paulo, São Paulo, Brazil
December 18, 1945

THE field produced by a point charge can be split in two parts:

(1) The attached field, $F^{\mu\nu}_{\text{at}}$;

(2) The radiated field, $F^{\mu\nu}_{\text{rad}}$.

We assume the total field to be the sum of the attached, the radiated and the external fields and that the sum of the attached and radiated fields gives the usual retarded field $F^{\mu\nu}_{\text{ret}}$ of the particle.

The radiated field behaves as an external field and reacts on the emitting charge. The attached field does not act on the generating charge but acts on other charges. Therefore, each charge is under the action of the retarded fields of the others, but only under the radiated part of its own field.

The particle of charge e and mass m has two kinds of momentum and energy: (1) the kinetic four vector

$$m\frac{dx^\mu}{ds} = G^\mu_{\text{kin}},$$

and (2) the four vector of the acceleration energy and momentum

$$G^\mu_{\text{ac}} = -\tfrac{2}{3}e^2\frac{d^2x^\mu}{ds^2}.$$

The four vector G^μ_{ac} arises from the interaction of the charge and its radiated field. Since the particle is under the action of the external field and its own radiated field, the equations of motion ought to be:

$$m\frac{d^2x^\mu}{ds^2} = \frac{e}{c}\frac{dx^\nu}{ds}(F^\mu_{\text{ext},\nu} + F^\mu_{\text{rad},\nu}). \tag{1}$$

If we apply the laws of conservation of energy and momentum we get:

$$\frac{d}{ds}(G^\mu_{\text{kin}} + G^\mu_{\text{ac}}) - \tfrac{2}{3}e^2\frac{dx^\mu}{ds}\frac{d^2x^\rho}{ds^2}\frac{d^2x_\rho}{ds^2} = \frac{e}{c}F^\mu_{\text{ext},\nu}\frac{dx^\nu}{ds}, \tag{2}$$

because the rate of loss of energy and momentum is

$$\tfrac{2}{3}e^2\frac{dx^\mu}{ds}\frac{d^2x^\rho}{ds^2}\frac{d^2x_\rho}{ds^2}$$

according to the well-known Larmor formulae. Comparing Eqs. (1) and (2) we see that:

$$\frac{e}{c}F^\mu_{\text{rad},\nu}\frac{dx^\nu}{ds} = -\frac{dG^\mu_{\text{ac}}}{ds} + \tfrac{2}{3}e^2\frac{dx^\mu}{ds}\frac{d^2x^\rho}{ds^2}\frac{d^2x_\rho}{ds^2}. \tag{3}$$

Equation (3) is satisfied by taking:

$$F^{\mu\nu}_{\text{rad}} = \tfrac{1}{2}(F^{\mu\nu}_{\text{ret}} - F^{\mu\nu}_{\text{adv}}). \tag{4}$$

From Eq. (4) it results that:

$$F^{\mu\nu}_{\text{at}} = \tfrac{1}{2}(F^{\mu\nu}_{\text{ret}} + F^{\mu\nu}_{\text{adv}}). \tag{5}$$

We define the stress-tensor $T^{\mu\nu}$ of the field in which there are n point charges in the following way:

$$4\pi T_\nu{}^{\mu'} = F^{\mu\rho}F_{\rho\nu} + \tfrac{1}{4}\delta_\nu{}^\mu(F^{\rho\sigma}F_{\rho\sigma})$$

$$-\Sigma_i F^{\mu\rho}_{i,\text{at}}F_{i,\text{at};\rho\nu} - \tfrac{1}{4}\delta_\nu{}^\mu(\Sigma_i F^{\rho\sigma}_{i,\text{at}}F_{i,\text{at};\rho\sigma}). \tag{6}$$

Equation (6) ensures the conservation of the energy and momentum of the system of the field and particles, and leads to a finite field energy which is not always positive. The energy and momentum of the field derived from the stress tensor $T^{\mu\nu}$ are the components of a four vector. A detailed exposition of the theory will be given elsewhere.

Our theory leads to the Lorentz-Dirac classical equations of motion and to a finite field energy without introducing any energies and momenta of non-electromagnetic nature besides the kinetic ones. There are no subtractions of infinite quantities.

The Radiation Field of a Point Electron

J. LEITE LOPES* AND MARIO SCHÖNBERG
Department of Physics, University of São Paulo, São Paulo, Brazil
December 18, 1944

THE radiation field of the electron has been defined in different ways. The most usual is to take it as the part of the retarded field of the charge that depends on the retarded acceleration and varies inversely with the retarded distance to the charge. Though this definition presents many obvious inconveniencies, it has the advantage of giving the right amount of the radiation emission. Dirac[1] has proposed the following definition of the radiation field of a point charge:

$$F^{\mu\nu}_{\text{rad},D} = F^{\mu\nu}_{\text{ret}} - F^{\mu\nu}_{\text{adv}}, \tag{1}$$

$F^{\mu\nu}$ being the antisymmetric tensor of the corresponding electromagnetic field: $\overset{\mu\nu}{F}_{\text{rad},D}$ the tensor of the radiation field, $\overset{\mu\nu}{F}_{\text{ret}}$ that of the retarded field and $\overset{\mu\nu}{F}_{\text{adv}}$ the tensor of the advanced field. Dirac's radiation field has many advantages: it is finite on the charge's world line, it is relativistically invariant and it is a solution of the homogeneous Maxwell equations. It has an inconvenience; it does not give the correct radiation losses. Indeed, it can be easily seen that:

$$e\overset{\mu}{F}_{\nu,\,\text{rad},D}\frac{dx^\nu}{ds}=\frac{4e^2}{3}\left[\frac{d^3x^\mu}{ds^3}+\frac{dx^\mu}{ds}\left(g_{\alpha\beta}\frac{d^2x^\alpha}{ds^2}\frac{d^2x^\beta}{ds^2}\right)\right], \tag{2}$$

if we take into account Dirac's expression of $\overset{\mu\nu}{F}_{\text{rad},D}$ on the charge's world line:

$$\overset{\mu\nu}{F}_{\text{rad},D}=\frac{4e}{3}\left[\frac{d^3x^\mu}{ds^3}\frac{dx^\nu}{ds}-\frac{d^3x^\nu}{ds^3}\frac{dx^\mu}{ds}\right], \tag{3}$$

$$x^0=ct,\quad x^1=x,\quad x^2=y,\quad x^3=z, \tag{4}$$

$$ds^2=g_{\alpha\beta}dx^\alpha dx^\beta=(dx^0)^2-(dx^1)^2-(dx^2)^2-(dx^3)^2. \tag{5}$$

The first term in the right-hand side of Eq. (2) is twice the rate of change of the four vector of acceleration energy and momentum and does not give rise to losses, the second term corresponds to twice the loss of energy and momentum by radiation given by the Larmor formulae. This shows that instead of Dirac's $\overset{\mu\nu}{F}_{\text{rad},D}$ it is better to take as the radiation field of a point charge $\overset{\mu\nu}{F}_{\text{rad}}$:

$$\overset{\mu\nu}{F}_{\text{rad}}=\tfrac{1}{2}(\overset{\mu\nu}{F}_{\text{ret}}-\overset{\mu\nu}{F}_{\text{adv}}). \tag{6}$$

Dirac's classical equations of motion[1] for a point charge can be put in the form:

$$m\frac{d^2x^\mu}{ds^2}=e\frac{dx^\nu}{ds}(\overset{\mu}{F}_{\nu,\,\text{ext}}+\overset{\mu}{F}_{\nu,\,\text{rad}}), \tag{7}$$

$\overset{\mu\nu}{F}_{\text{ext}}$ being the tensor of the external field, arising from other charges. Equation (7) shows that the charged particle is under the action of the total field and its own radiation field defined according to Eq. (6), which should be reasonably expected since only the radiation field which gets separated from the particle can react on it.

A consistent theory of the radiation field of the electron that will be given elsewhere by one of us shows that all the difficulties of the classical theory of the electron can be overcome without any subtraction processes of the kind involved in the theories of Dirac and Pryce.[2]

[1] P. A. M. Dirac, Proc. Roy. Soc. **167**, 148 (1938).
[2] M. H. L. Pryce, Proc. Roy. Soc. **168**, 389 (1938).
* Now at Princeton University with a fellowship of the U. S. Department of State under the provisions of the Convention for the Promotion of Inter-American Cultural Relations. This work was done in 1943 when J. Leite Lopes was at the University of S. Paulo with a fellowship of the Zerrener Foundation.